NASA Contractor Report 172009

National Aeronautics and Space Administration (NASA)/ American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program--1987

Volume 2

William B. Jones, Jr., Editor
Texas A&M University
College Station, Texas

Stanley H. Goldstein, Editor
University Programs Office
Lyndon B. Johnson Space Center
Houston, Texas

Grant NGT-44-001-800

November 1987

National Aeronautics and Space Administration

Lyndon B. Johnson Space Center
Houston, Texas

PREFACE

The 1987 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/-American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by Texas A&M University and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began in 1965 at JSC and in 1964 nationally, are

a. to further the professional knowledge of qualified engineering and science faculty members;

b. to stimulate an exchange of ideas between participants and NASA;

c. to enrich and refresh the research and teaching activities of participants' institutions; and

d. to contribute to the research objectives of the NASA Centers.

Each faculty fellow spent 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1987. Volume 1 contains sections 1 through 18, and volume 2 contains sections 19 through 35.

CONTENTS

A GENERAL ARCHITECTURE FOR INTELLIGENT TRAINING SYSTEMS

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	R. Bowen Loftin, Ph.D.
Academic Rank:	Associate Professor
University & Department	University of Houston-Downtown Department of Natural Sciences One Main Street Houston, Texas 77002
NASA/JSC	
Directorate:	Mission Support
Division:	Mission Planning and Analysis
Branch:	Technology Development and Applications
Section:	Artificial Intelligence
JSC Colleague:	Robert T. Savely
Date:	August 14, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

A preliminary design of a general architecture for autonomous intelligent training systems has been developed. The architecture integrates expert system technology with teaching/training methodologies to permit the production of systems suitable for use by NASA, other government agencies, industry, and academia in the training of personnel for the performance of complex, mission-critical tasks. The proposed architecture consists of five elements: a user interface, a domain expert, a training session manager, a trainee model, and a training scenario generator. A user interface permits the trainee to access data and perform actions as he would in the task environment. The interface also provides the trainee with information on the current training environment and with on-line help. The domain expert contains the rules and procedural knowledge needed to carry out the task. It also includes "mal-rules" which permit the identification of common errors on the part of the trainee. The training session manager (TSM) examines the actions of the trainee and compares them with the actions of the domain expert. A unique feature of the TSM is its ability to permit the trainee the freedom to follow any valid path between two stages of the task for which he is being trained. Following each trainee action, evaluative assertions are made by the TSM and used to update the trainee model. A trainee model is developed for each individual using the system. The model includes a history of the trainee's interactions with the intelligent training system and provides evaluative data on the trainee's current skill level. A training scenario generator designs appropriate training exercises for each trainee based on the trainee model and the training goal(s). The design of this architecture has been guided and its efficacy tested through the development of a system for use by Mission Control Center Flight Dynamics Officers in training to perform Payload-Assist Module Deploys from the orbiter.

INTRODUCTION

Government and industry must maintain a large effort in the training of their personnel: new personnel must be trained to perform the task(s) which they were hired to perform, continuing personnel must be trained to upgrade or update their ability to perform assigned tasks, and continuing personnel must be trained to tackle new tasks. Within NASA a great number of training methodologies are employed, singly or in concert. These methods include training manuals, formal classes, procedural computer programs, simulations, and on-the-job training. The latter method is particularly effective in complex tasks where a great deal of independence is granted to the task performer. Of course, this training method is also the most expensive and may be impractical when there are many trainees and few experienced personnel to conduct on-the-job training.

This report describes an effort to design an architecture for an entire generation of autonomous training systems which integrate many of the features of "traditional" training programs with expert system technology. The ultimate goal of this program is a software development environment which would permit those responsible for training in government, industry, and academia to develop intelligent computer-aided training (ICAT) systems for specific tasks. These ICAT systems would operate autonomously and would provide, for the trainee, much of the same experience that could be gained from the best on-the-job training. By integrating domain expertise with a knowledge of appropriate training methods, an ICAT session is intended to duplicate, as closely as possible, the trainee undergoing on-the-job training in the task environment, benefiting from the full attention of a task expert who is also an expert trainer. Thus, the philosophy of the ICAT system is to emulate, to the extent possible, the behavior of an experienced individual devoting his full time and attention to the training of a novice--proposing challenging training scenarios, monitoring and evaluating the actions of the trainee, providing meaningful comments in response to trainee errors, responding to trainee requests for information and hints (if appropriate), and remembering the strengths and weaknesses

displayed by the trainee so that appropriate future exercises can be designed.

BACKGROUND

Since the 1970's a number of academic and industrial researchers have explored the application of artificial intelligence concepts to the task of teaching a variety of subjects (e.g., geometry, computer programming, medical diagnosis, and electronic troubleshooting). A body of literature is now extant on student models and teaching/tutoring methodologies adapted to intelligent tutoring systems in the academic environment[1]. The earliest published reports which suggested the applications of artificial intelligence concepts to teaching tasks appeared in the early 1970's.[2,3] Hartley and Sleeman[3] actually proposed an architecture for an intelligent tutoring system. However, it is interesting to note that, in the fourteen years which have passed since the appearance of the Hartley and Sleeman proposal, no agreement has been reached among researchers on a general architecture for intelligent tutoring systems.[4] Nonetheless, a study of the literature on intelligent tutoring systems is an essential starting point for the development of the elements of an intelligent training system.

Among the more notable intelligent tutoring systems reported to date are SOPHIE[5], PROUST[6] and the LISP Tutor[7]. The first of these systems, SOPHIE, was developed in response to a U.S. Air Force interest in a computer-based training course in electronic troubleshooting. SOPHIE contains three major components: an electronics expert with a general knowledge of electronic circuits, together with detailed knowledge about a particular type of circuit (in SOPHIE this was an IP-28 regulated power supply); a coach which examines student inputs and decides if it is appropriate to stop the student and offer advice; and a troubleshooting expert that uses the electronics expert to determine which possible measurements are most useful in a particular context. Three versions of SOPHIE were produced and used for a time but none was ever viewed as a "finished" product. One of the major lacks of the SOPHIE systems was a user model. It is interesting to note that the development of a natural language interface for SOPHIE represented a large portion of the total task.

PROUST and the LISP Tutor are two well-known intelligent tutoring systems that have left the laboratory and found wider applications. PROUST (and its offspring, Micro-PROUST) serves as a "debugger" for finding nonsyntactical errors in Pascal programs written by student programmers. The developers of PROUST claim that it is capable of finding all of the bugs in at least seventy percent of the "moderately complex" programming assignments that its examines. PROUST contains an expert Pascal programmer that can write "good" programs for the assignments given to students. Bugs are found by matching the expert's program with that of the student; mismatches are identified as "bugs" in the student program. This ability is contained in the PROUST "bug rule" component. After finding a bug, PROUST provides an English-language description of the bug to the student, enabling the student to correct his error. The system cannot handle student programs that depart radically from the programming "style" of the expert. The LISP Tutor is currently used to teach the introductory Lisp course offered at Carnegie-Mellon University. This system is based on the ACT (historically, Adaptive Control of Thought) theory and consists of four elements: a structured editor which serves as an interface to the system for students, an expert Lisp programmer that provides an "ideal" solution to a programming problem, a bug catalog that contains errors made by novice programmers, and a tutoring component that provides both immediate feedback and guidance to the student. Evaluations of the LISP Tutor show that it can achieve results similar to those obtained by human tutors. One of its primary features is its enforcement of what its authors regard as a "good" programming style.

TRAINING VERSUS TUTORING

The ICAT architecture has been developed with a clear understanding that training is not the same as teaching or tutoring.[8] The NASA training environment differs in many ways from an academic teaching environment These differences are important in the design of an architecture for an intelligent training system:

a. Assigned tasks are often mission-critical, placing the responsibility for lives and property in the hands of those who have been trained.

b. Personnel already have significant academic and practical experience to bear on their assigned

c. Trainees make use of a wide variety of training techniques, ranging from the study of comprehensive training manuals to simulations to actual on-the-job training under the supervision of more experienced

d. Many of the tasks offer considerable freedom in the exact manner in which they may be accomplished.

Those undergoing training for complex, mission-critical tasks are usually well aware of the importance of their job and the probable consequences of failure. While students are often motivated by the fear of receiving a low grade, trainees know that human lives and/or expensive equipment may depend on their skill in performing assigned tasks. This means that trainees may be highly motivated, but it also imposes on the trainer the responsibility for the accuracy of the training content (i.e., verification of the domain expertise encoded in the system) and the ability of the trainer to correctly evaluate trainee actions. The ICAT system is intended, not to impart basic knowledge, but to aid the trainee in developing skills for which he already has the basic or "theoretical" knowledge. In short, this training system is designed to help a trainee put into practice that which he already intellectually understands. The system must take into account the type of training that both precedes and follows, building on the knowledge gained from training manuals and rule books while preparing the trainee for and complementing the on-the-job training which will follow. Perhaps most critical of all, trainees must be allowed to carry out an assigned task by any valid means. Such flexibility is essential so that trainees are able to retain and even hone an independence of thought and develop confidence in their ability to respond to problems, even problems which they have never encountered and which their trainers never anticipated.

SYSTEM DESIGN

The ICAT system architecture is modular and consists of five basic components:

1. A user interface that permits the trainee to access the same information available to him in the the task environment and serves as a means for the trainee to assert actions and communicate with the intelligent training system
2. A domain expert which can carry out the task using the same information that is available to the trainee and which also contains a list of "mal-rules" (explicitly identified errors that novice trainees commonly make).
3. A training session manager (TSM) which examines the assertions made by the domain expert (of both correct and incorrect actions in a particular context) and by the trainee. Evaluative assertions are made following each trainee action. In addition, guidance can be provided to the trainee if appropriate for his skill level.
4. A trainee model which contains a history of the individual trainee's interactions with the system together with summary evaluative data.
5. A training scenario generator that designs increasingly-complex training exercises based on the current skill level contained in the trainee's model and on any weaknesses or deficiencies that the trainee has exhibited in previous interactions.

Figure 1 contains a schematic diagram of the ICAT system. Note that provision is made for the user to interact with the system in two distinct ways and that a supervisor may also query the system for evaluative data on each trainee. The blackboard serves as a common "factbase" for all five system components. With the exception of the trainee model, each component makes assertions to the blackboard, and the rule-based components look to the blackboard for facts against which to pattern match the left-hand sides of their rules.

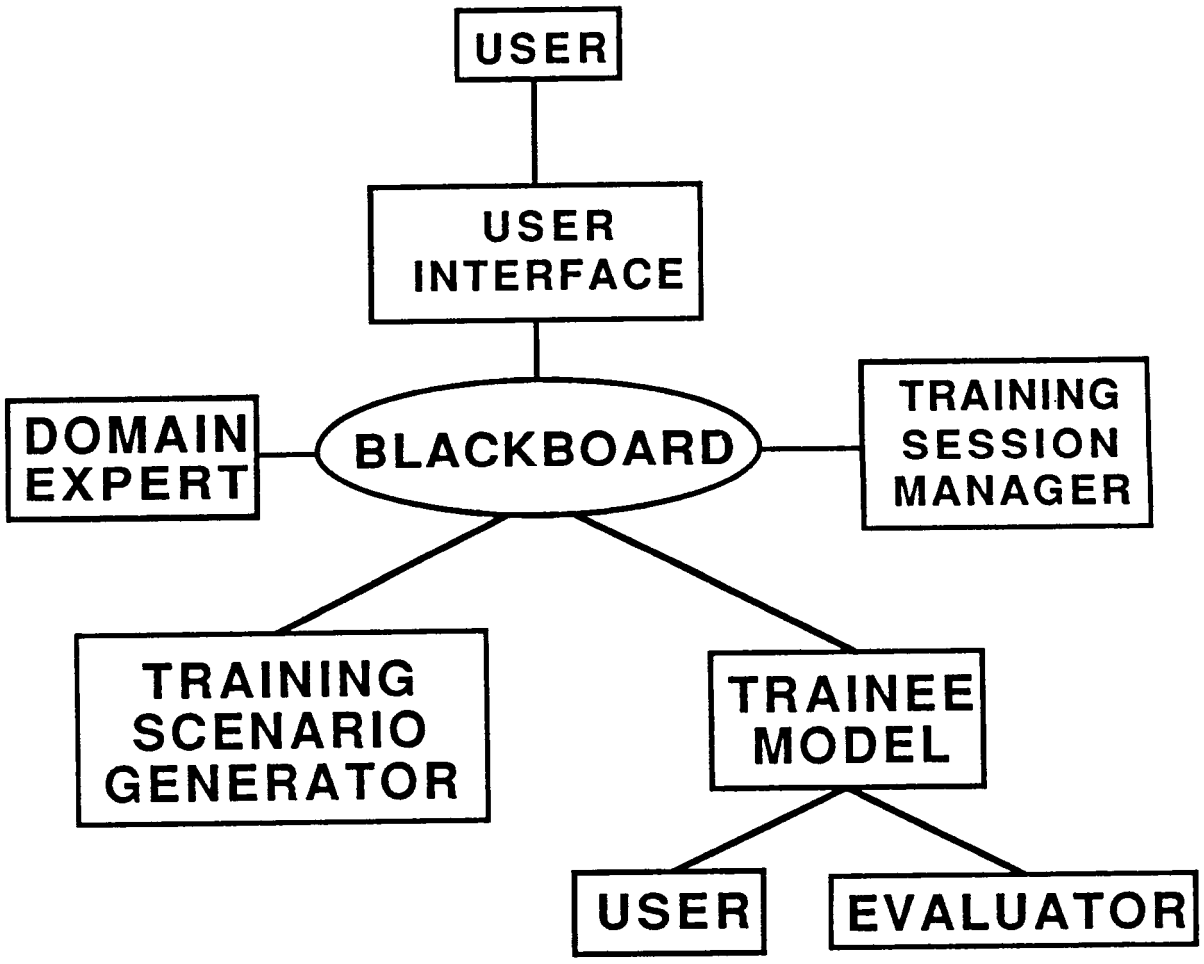


FIGURE 1. ICAT ARCHITECTURE

User Interface

The primary factor influencing the interface design is fidelity to the task environment. To avoid negative training, it is essential that the functionality and, to the extent possible, the actual appearance of the training environment duplicate that in which the task is performed. The interaction of the user interface with the other components of the ICAT system can be standardized and transported to any task environment; however, the user interface, as viewed by the trainee, must be developed for each task environment. The training of different tasks in the same environment may be able to utilize the same user interface. As a part of this project, a user interface was designed for a console position in the Mission Control Center at NASA/Johnson Space Center. The details of that interface are described elsewhere.[9]

Domain Expert

The domain expert is a "traditional" expert system in that it contains if-then rules which access data describing the task environment and is capable of executing the task and arriving at the correct "answers" or performing the correct actions. In addition to "knowing" the right way to carry out the task, the domain expert also contains knowledge of the typical errors that are made by novices. In this way, the ICAT system can not only detect an erroneous action made by a trainee, but also, through these so-called "mal-rules", it can diagnose the nature of the error and provide an error message to the trainee specifically designed to inform the trainee about the exact error made and correct the misconception or lack of knowledge that led to the commission of that error. Another of the interesting features of the ICAT system is its continual awareness of the environment (the external constraints dictated by the training exercise) and the context of the exercise. Rather than having the domain expert generate a complete and correct set of actions to accomplish the task, only those actions which are germane to the current context are asserted. In this way the expert "adapts" to alternate, but correct, paths that the trainee might choose to follow. Figure 2 shows schematically how the domain expert operates. This strategy was adopted because the human experts that perform complex tasks often recognize that many steps in the process may

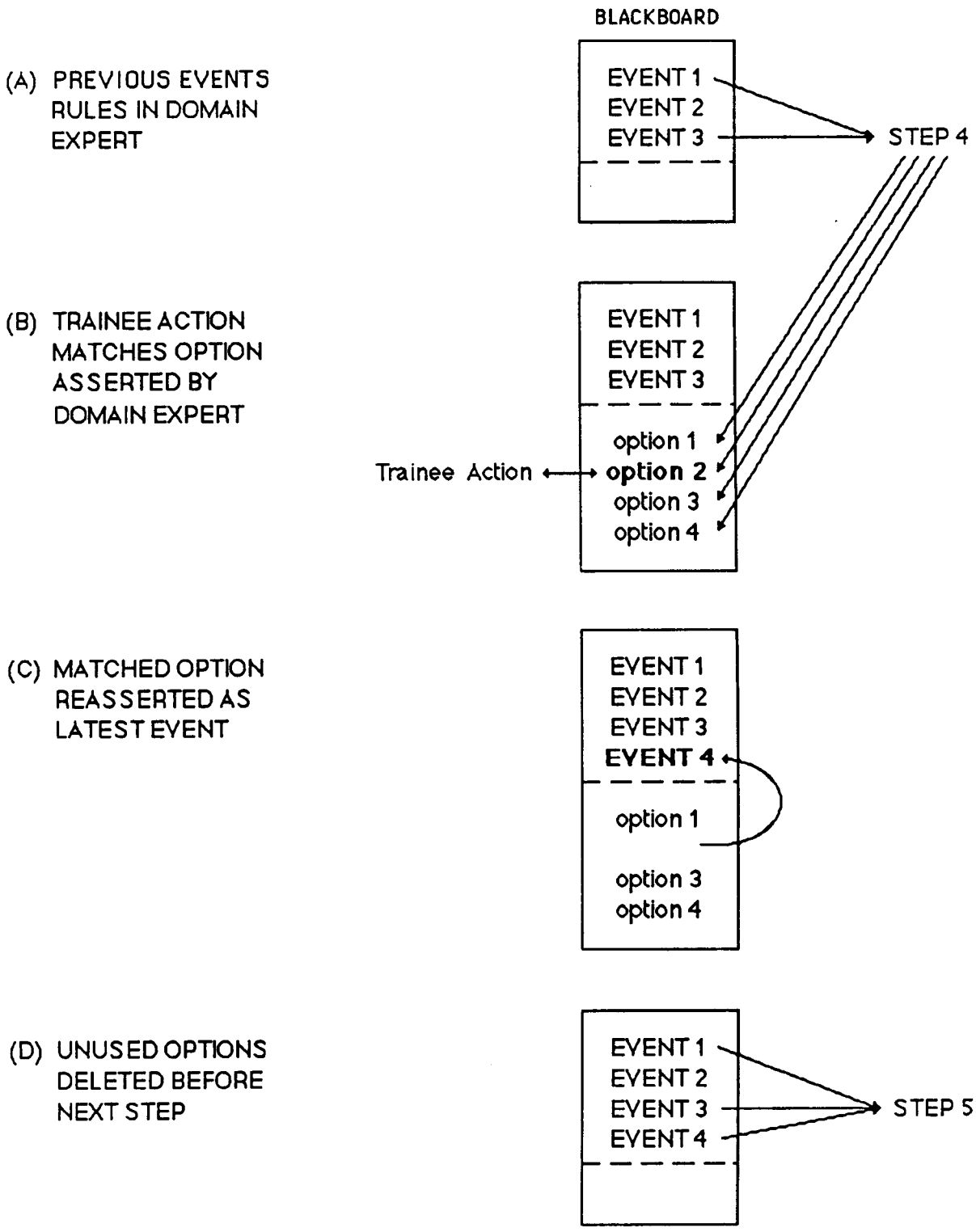


FIGURE 2. OPERATION OF DOMAIN EXPERT

be accomplished by two or more equally valid sequences of actions. To grant freedom of choice to the trainee and to encourage independence on his part, this type of flexibility in the ICAT system was deemed essential.

Training Session Manager

The training session manager is dedicated principally to error-handling. Its rules compare the assertions of the domain expert with those of the trainee to detect errors. Subsequently, the domain expert asserts facts that allow the TSM to write appropriate error messages to the trainee through the user interface. In addition, the TSM is sensitive to the skill level of the trainee as represented by the trainee model. As a result, the detail and "tone" of error messages is chosen to match the current trainee. For example, an error made by a first time user of the training system may require a verbose explanation so that the system can be certain that the trainee will have all of the knowledge and concepts needed to proceed. On the other hand, an experienced trainee may have momentarily forgotten a particular procedure or may have "lost his place". In this latter case a terse error message would be adequate to allow the trainee to resume the exercise. The TSM also encodes all trainee actions, both correct and incorrect, and passes them to the trainee model.

Trainee Model

Successful intelligent tutors incorporate student models to aid in error diagnosis and to guide the student's progress through the tutor's curriculum.[10] The trainee model in the ICAT system stores assertions made by the TSM as a result of trainee actions. Thus, at its most fundamental level, the trainee model contains, for the current training session, a complete record of the correct and incorrect actions taken by the trainee. At the conclusion of each training session, the model updates a training summary which contains information about the trainee's progress, such as a skill level designator, number of sessions completed, number of errors made (by error type and session), and the time taken to complete sessions. After completing a session, the trainee can obtain a report of that session which contains a comprehensive list of correct and incorrect actions together with an evaluative commentary. A supervisor can

access each trainee's model to obtain this same report or to obtain summary data, at a higher level, on the trainee's progress. Finally, the training scenario generator uses the trainee model to produce new training exercises.

Training Scenario Generator

The training scenario generator relies upon a database of task "problems" to structure unique exercises for a trainee each time he interacts with the system. The initial exercises provided to a new trainee can be based on variants of the purely nominal task with no time constraints, distractions or "problems". Once the trainee has demonstrated an acceptable level of competence with the nominal task, the generator draws upon its database to insert selected problems into the training environment. In addition, time constraints are "tightened" as the trainee gains more experience and distractions of a form appropriate for the task environment are presented at "inconvenient" points during the task. The generator also examines the trainee model for particular types of errors committed by the trainee in previous (and the current) sessions. The trainee is then given the opportunity to demonstrate that he will not make that error again. Ultimately, the trainee is presented with exercises which embody the most difficult problems together with time constraints and distractions comparable to those encountered during the completion of an experienced person in the actual task environment.

SYSTEM INTEGRATION

The ICAT system which has been built as a prototype for this project is currently operational on a Symbolics 3600 series Lisp machine. The user interface and trainee model are written in common Lisp while the rules of the domain expert, TSM, and the training scenario generator are written in ART 3.0. This prototype system will ultimately be delivered to the users in a Unix workstation environment. To accomplish this delivery, the ART rules were written to facilitate translation into to CLIPS[11] and the Lisp code will be converted into C.

CONCLUSIONS

The prototype ICAT system has, so far, proven to be a potentially valuable addition to the training tools available for training Flight Dynamics Officers in shuttle ground control. The authors are convinced that the basic structure of the ICAT system described here can be extended to form a general architecture for intelligent training systems for training flight controllers and crew members in the performance of complex, mission-critical tasks. It may ultimately be effective in training personnel for a wide variety of tasks in governmental, academic, and industrial settings.

REFERENCES

1. See, for example, Sleeman, D. and Brown, J.S. (eds.), Intelligent Tutoring Systems (London: Academic Press, 1982) and Yazdani, M. "Intelligent Tutoring Systems Survey," Artificial Intelligence Review **1**, 43 (1986).

2. Carbonell, J.R. "AI in CAI: An Artificial Intelligence Approach to CAI," IEEE Transactions on Man-Machine Systems **11**(4), 190 (1970).

3. Hartley, J.R. and Sleeman, D.H., "Towards Intelligent Teaching Systems," International Journal of Man-Machine Studies **5**, 215 (1973).

4. Yazdani, M. "Intelligent Tutoring Systems Survey," Artificial Intelligence Review **1**, 43 (1986).

5. Brown, J.S., Burton, R.R., and de Kleer, J., "Pedagogical, Natural Language and Knowledge Engineering Techniques in SOPHIE I, II, and III," in Sleeman, D. and Brown, J.S., (eds.), Intelligent Tutoring Systems (London: Academic Press, 1982), p. 227.

6. Johnson, W.L. and Soloway, E. "PROUST, " Byte **10** (4), 179 (April, 1985).

7. Anderson, J.R., Boyle, C.F., and Reiser, B.J., "Intelligent Tutoring Systems," Science **228**, 456 (1985) and Anderson, J.R. and Reiser, B.J., "The LISP Tutor," Byte **10**(4), 159 (April, 1985).

8. Harmon, P. "Intelligent Job Aids: How AI Will Change Training in the Next Five Years," in Kearsley, G., ed., Artificial Intelligence and Instruction: Applications and Methods (Reading, MA: Addison Wesley Publishing Co., 1987).

9. Loftin, R.B., Wang, L., Baffes, P., and Rua, M., "An Intelligent Computer-Aided Training System for Payload-Assist Module Deploys," Proceedings of the First Annual Workshop on Space Operations, Automation & Robotics (SOAR '87), NASA/Johnson Space Center, Houston, TX, August 4-6, 1987.

10. See, for example, a number of papers on student models in Sleeman, D. and Brown, J.S., (eds.), Intelligent Tutoring Systems (London: Academic Press, 1982)

11. "CLIPS" is an acronym for "C-Language Integrated Production System" and was developed by the Artificial Intelligence Section, Mail Code FM72, NASA/Johnson Space Center, Houston, TX 77058. Its advantages as a delivery vehicle for expert systems are discussed in Giarratano, J., Culbert, C., Riley, G., and Savely, R.T., "A Solution of the Expert System Delivery Problem," submitted for publication in IEEE Expert. For additional information on CLIPS, write to the AI Section at NASA/JSC.

DYNAMICS, CONTROL AND SENSOR ISSUES PERTINENT TO ROBOTIC HANDS FOR THE EVA RETRIEVER SYSTEM

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Robert A. McLauchlan, Ph.D.,P.E.
Academic Rank:	Associate Professor
University & Department:	Texas A & I University Civil & Mechanical Engineering Department Kingsville, Texas 78363
NASA/JSC	
Directorate:	Engineering
Division:	Crew & Thermal Systems
Branch:	Engineering Crew (EC2, EC5)
JSC Colleagues:	Cliff Hess, EC5 Paul Marshall, EC2
Date:	September 3, 1987
Contract:	NGT 44-001-800

ABSTRACT

Basic dynamics, sensor, control and related artificial intelligence issues pertinent to smart robotic hands for the Extra Vehicular Activity (EVA) Retriever system are summarized and discussed. These smart hands are to be used as end effectors on arms attached to manned maneuvering units (MMU). The Retriever robotic systems comprised of MMU, arm and smart hands, are being developed to aid crewmen in the performance of routine EVA tasks including tool, object, etc. retrieval. The ultimate goal of this work is to enhance the effectiveness of EVA crewmen.

INTRODUCTION

Overview

Manned Extra Vehicular Activity (EVA) is an important part of planned orbital Space Station development and operations as, e.g.,

* Initial assembly, on-going maintenance of the Space Station.
* Maintenance and servicing - as per refueling and repair of satellites, platforms, and free-flyers.

Manned EVA is also available to be used as backup to telerobotic systems and for unplanned or unique operations.

Here EVA safety needs require a man-in-a-suit retrieval capability. More routine needs for such things as tool and other item retrieval, and assistance with workpiece/object - handling and maneuverability during manned-maintenance and construction tasks, also require at least a low-level, dextrous and intelligent robotic assistant capability.

Retriever Rationale and Background

More specifically, during EVA operations the potential exists for an EVA crewman or piece of equipment to separate from or inadvertently be launched away from the Space Station. Equipment failure or procedural error may cause such separation. Proposed Space Station EVA procedures require that there be at least two crewmen present for any EVA. However, it may not be practical because of safety considerations or time constraints for the second crewman to (a) Properly equip himself, (b) Leave the Space Station, and (c) Retrieve the lost equipment or rescue the first crewman who may be helpless.

Figure 1 from [Ref. 1] plots range versus range rate of separation for an object which is drifting free from the Space Station. The initial velocity for the drifting object is 2 ft/sec in the case depicted. Here estimates of the position and drift rate of the object are indicated at the discrete data points shown.

Table 1 summarizes these estimates of position and drift rate as functions of time. It indicates that at ten minutes from separation the object will already be 1400 feet away and have a rate of separation of 3.2 ft/sec. These values are 2350 ft and 5.3 ft/sec ,respectively at fifteen minutes from separation. The results summarized in

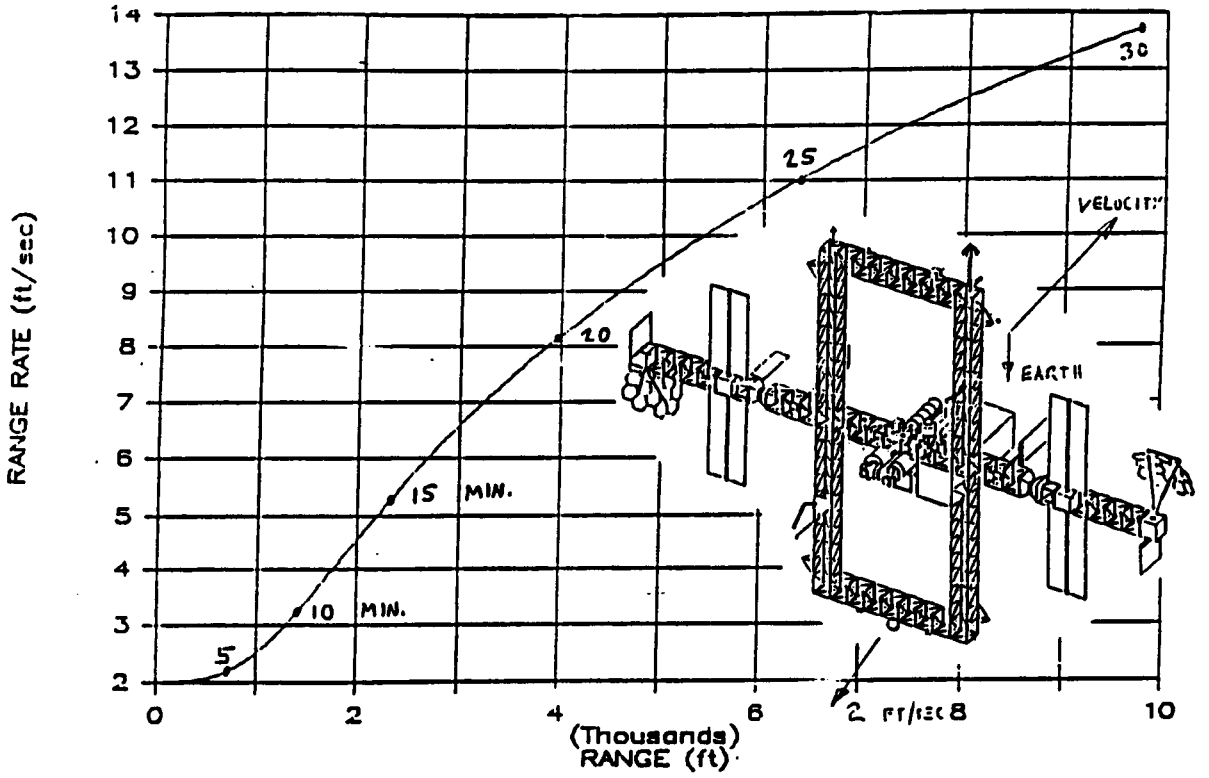
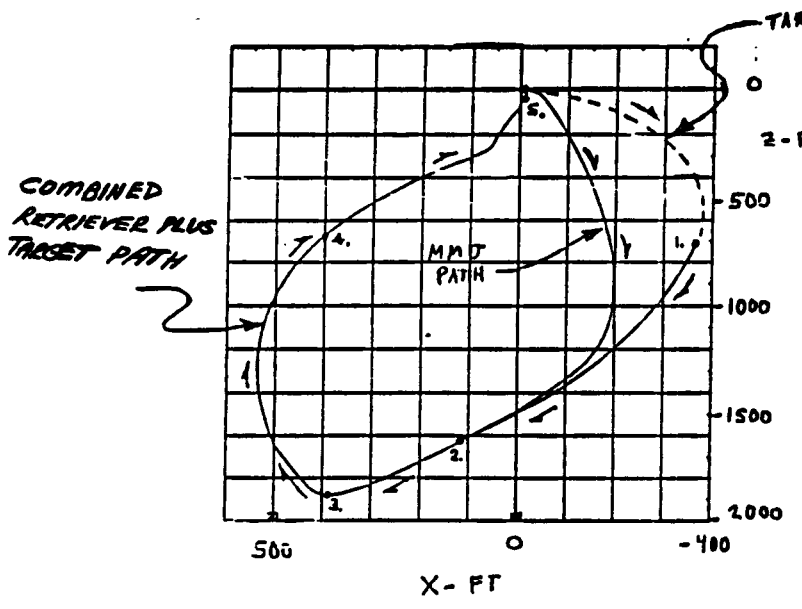


FIGURE 1 RANGE VERSUS RANGE RATE OF SEPARATION FOR AN OBJECT DRIFTING FREE FROM THE SPACE STATION



POINT NUMBER I	DRIFT RATE R (FPS)	TIME T (MIN)
1.	1.42	10.0
2.	2.65	17.2
3.	-.17	19.1
4.	-2.98	25.2
5.	-.26	31.0

SUMMARY OF RETRIEVAL STAGES

- APPROACH T_1 to T_2
- CAPTURE T_2 to T_{3-}
- GRASP T_{3-} to T_{3+}
- ADJUSTMENT AND RETURN T_{3+} to T_5

FIGURE 2 REPRESENTATIVE TRAJECTORY FOR RETRIEVAL SCENARIO

Table 1. Estimated Range and Range Rate as Functions of Time

Time T, Min	Range R, Kft	Range Rate R, ft/sec
5	0.07	2.2
10	1.40	3.2
15	2.35	5.3
20	3.96	8.2
25	6.39	11.0
30	9.67	13.7

Table 1 thus emphasize that time is extremely critical in initiating any retrieval operation. This is especially true if elapsed time and fuel costs are to be routinely minimized and indeed successful retrieval be assured.

Figure 2 depicts a representative trajectory for a retrieval scenario with an astronaut piloted Manned Maneuvering Unit (MMU) chasing a target. Here drift rate and time at five discrete times in the range from 10 to 31 minutes. Figure 2 also illustrates the urgency of initiating rescue as quickly as possible after separation.

JSC has developed the concept of the EVA Retriever, which is a highly autonomous, free-flying robot or robotic system. In use it will be on standby during EVA operations to provide the immediate chase, capture, and return capability required for adrift crewmen or station equipment. The Retriever will be used to fill the Space Station requirement which has been authorized by Space Station Control Board Directive BB000169A [Ref. 1]. The conceptual design of the EVA Retriever is depicted in Figure 3. By way of summary, Ref.1 indicates that the Retriever is:

* Intended to be highly autonomous
* Capable of locating and tracking the target
* Able to plan a path to the target and to avoid obstacles encountered during the chase, etc. segment
* Able to grapple - i.e., capture and grasp, etc. the target through the use of dextrous arms and hands
* Intended to interface with the Space Station MMU - thereby providing commonality with the crew interface
* Activated and supervised by voice command.

Table 2 gives a descriptive summary of the attributes of the smart hand plus arm plus MMU - EVA Retriever robotic system. Here 14 attributes ranging from principal function, environment, control architecture, etc. through sensors, computer processing, construction, and power are considered. These autonomous EVA Retriever attributes are somewhat different from those of a space telerobotic system. It should be noted that this system is directed toward the true leveraging of an astronaut crewman during EVA. Here an analogy can be drawn to a shepherd with his dog(s) working a flock. A telerobotic system on the other hand will occupy at least one astronaut crewman per system at a maximum work efficiency of 50 % of one EVA crewman.

Research Work Description

The work reported herein is part of an overall effort to investigate the intelligent control of robotic hands on the EVA Retriever. It has involved the exploration of basic dynamics, sensor,

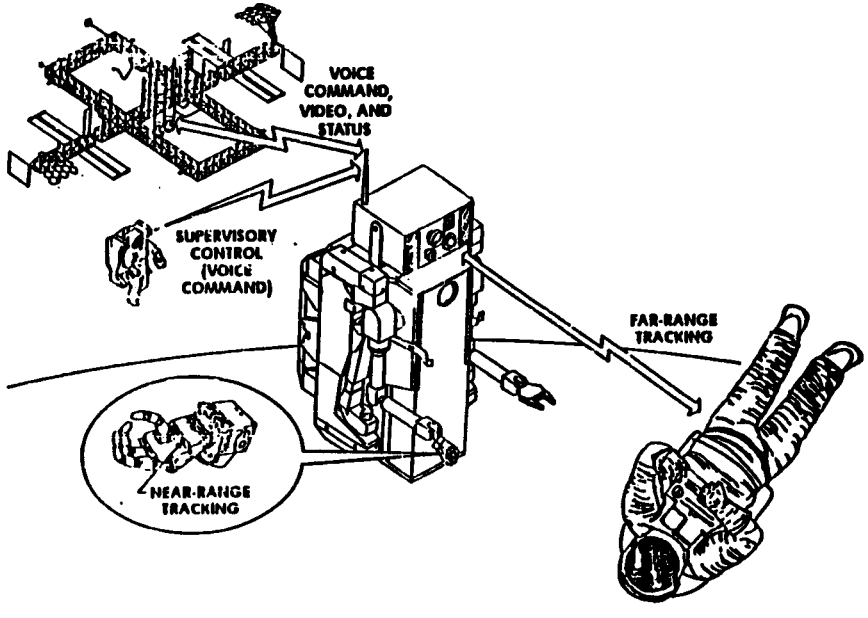


FIGURE 3 CONCEPTUAL DESIGN OF THE EVA RETRIEVER

TABLE 2, SMART HAND, ARM, FMU RETRIEVER SYSTEM ATTRIBUTES

Attribute	Description
1. Principal Function	Repetitive and/or Hazardous EVA Operations: Tool, Crewmen Retrieval.
2. Environment	Complex, Time Limited, Often Hazardous.
3. Primary Control Architecture, Parameters	Parallel Time, Hierarchical Task Accomplishment; Deep Knowledge: Position, Rate, and Force.
4. Operating Speed	Task Dependent, Subject to Human, Robot, and Equipment Safety -- As Fast as Possible.
5. Load Range	5 to 8 Lbm * Hand-Fingers 8 to 20 Lbm * Finger Wrap + wrist 20 to 450 Lbm * Arm + Finger-lock. * At 1g Acceleration.
6. Kinematics	Arm/Hand Rotary, Prismatic (and Spherical) Joints; Various DOF.
7. End-Effectors	General Purpose Grippers/ Fingers
8. Compliance	Hand - Relatively Flexible, Backdriveable, Some Backlash. Arm/Wrist - Relatively Stiff, Ideally Backdriveable, Minimal Backlash.
9. Actuators	Hand - Usually Local or Close to Joint. Arm/Wrist - Local or Close to Joint.
10. Transmission	Arm - Gear Box Drive; Hand - Torque Tubes, Kevlar Cable and Pulley Tendons, FourBar and Related Linkages.
11. Sensors	Vision - Nearfield/Farfield; Tactile - Proximity and Contact/Touch. Joint Torque, Position Sensors as Close to Joint as Possible.
12. Computer Processing Power	Parallel - Transputer/Multi-processor; Significant Pre-Processing of Signals at the Sensor Array(s); Potential Use of Multilayer Neural Net Arrays.
13. Construction	Possibly Composites; Space Grade Aluminum, Steel, and Titanium as Appropriate.
14. Power	Electric; Electromechanical, Piezoelectric.

control, and related Artificial Intelligence issues pertinent to robotic hands. These hands are to be used as end effectors on arms attached to Manned Maneuvering Units (MMU). The Retriever robotic systems comprised of MMU, arm and smart hands, are being developed to aid crewmen in the performance of routine EVA tasks including tool, object, etc. retrieval. The ultimate goal of this work is to enhance the effectiveness of EVA crewmen.

BASIC ISSUES

Smart, Dextrous Robotic Hands

At this point it is important to consider the requirements (or desirable characteristics of typical smart robotic hands for the NASA/JSC EVA Retriever. Figure 4 shows a later generation version (4 articulated fingers - 3 fingers plus a thumb) which the Crew and Thermal Systems Division (CTSD) at NASA/JSC is using. As far as dextrous hands are concerned, this is perhaps the best hand available in the world today. However, its pneumatic servo actuation system is impractical as far as space operations of the EVA Retriever are concerned. CTSD at NASA/JSC has also or will shortly have the Salisbury/JPL hand, the Jameson/Stanford hand, and the Grubbs/CTSD hand. These latter hands each have three articulated fingers. Along with the four-fingered Utah/MIT hand, they are being used by CTSD to explore basic issues relevant the dynamics, sensing, and control of (a) the smart hands per se, and (b) the integrated smart hand plus articulated arm plus MMU - comprising the EVA Retriever robotic system. In addition, related Artificial Intelligence (AI) and computer issues are being investigated.

Table 2 as discussed previously summarizes the overall characteristics of the EVA Retriever. here it is seen that the dextrous hand/wrist/arm system is of great importance to EVA Retriever success. It is important to note that the hand-wrist-arm combinations cover the load range typically encountered by a human arm in a 1 g acceleration field.

Figure 5 shows the six different types of prehension which comprise human hand grasping. These prehensions characterize the basic object grasp or gripping functions which humans use everyday. Here the six grasps can be interpreted as basic operations for smart robotic hands as well.

Table 3. Summary of Manipulation Functions Performed by the Human Hand

1. Trigger Grip
2. Flipping a Switch
3. Transferring Pipe to Grip
4. Using Cutters/Shears
5. Screwing a Pen
6. Rolling a Cylindrical Object
7. Transferring a Pen
8. Typewriting/Using a Keyboard
9. Writing With a Pen

Nine basic manipulation functions performed by the human hand

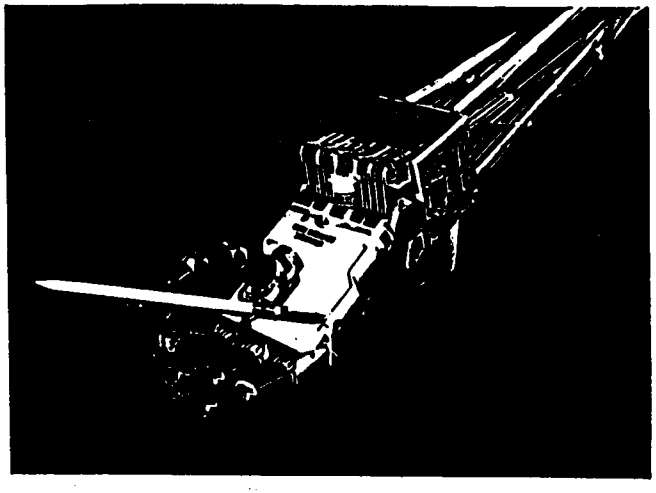


FIGURE 4, Later Generation Version of the UTAH/MIT Dextrous Hand Used by CTSD at NASA/JSC.

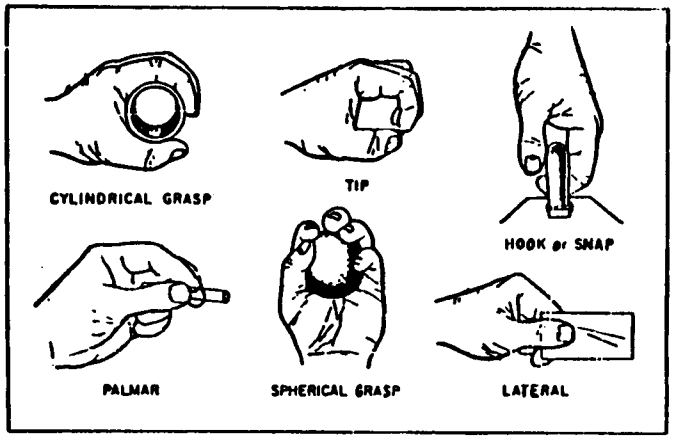


FIGURE 5, SUMMARY OF THE SIX DIFFERENT TYPES OF PREHENSION COMPRISING HUMAN-HAND GRASPING

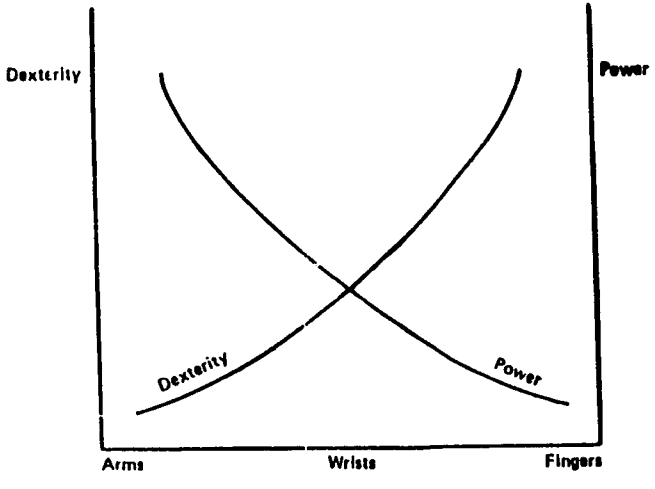


FIGURE 6, REPRESENTATIVE TRADE-OFF BETWEEN ARM, WRIST AND FINGER MANIPULATION. NOTE THAT THE ABSCISSA AXIS HAS UNITS OF APPLIED FORCE ($F_{fingers}$ & F_{wrist} & F_{arms})

Table 4. Summary of the Approximate Relations Between the Number of Fingers and Joints, and the Functions of the Hand/End Effector.

Hand Finger Configurations		Functions		
		Grasping	Shape Accommodation	Manipulation
Number	Type			
2	Rigid	Yes	None	None
2	Articulate	Yes	Yes-Some	None
3	Rigid	Yes	Yes-Some	None
3	Articulate	Yes	Yes-Some	Yes-Some
4	Rigid	Yes	Yes-Some	Yes-Some
4	Articulate	Yes	Yes	Yes
5	Rigid	Yes	Yes	Yes-Some
5	Articulate	Yes	Yes	Yes

are summarized in Table 3. Of these manipulation functions, the 'lower level' ones: (1) Trigger grip, (2) Flipping a switch, (3) Rolling a cylindrical object, (4) Transferring pipe to grip, (5) Screwing a pen (or a bolt-nut combination), (6) Using cutters /shears -- are considered by the author as necessary for the EVA Retriever. This is especially true if the Retriever is to be used for assistance with a workpiece/in object handling and thereby, truly enhance the effectiveness of EVA crewmen.

Humans perform grasp and manipulate functions on a routine, everyday basis with their five finger (four fingers plus an opposable thumb) hands. Table 4 indicates that the six grasp and at least six of the manipulate functions can be performed with three or more fingered hands. Here each finger in multiple (three to four revolute) jointed. Figure 6 gives representative plots of dexterity and power as functions of the type of element (finger/hand, wrist, arm) combination or force level applied. These dexterity and power curves are composites of the maximum value portions of the curves for (a) Hand-fingers, (b) Finger wrap plus wrist, and (c) Arm plus finger-lock combinations. The crossover points occur at nominal values of 8, 20, and 250 Lbf (8, 20, and 250 Lbm in 1 g acceleration field). As expected, hand-finger combinations exhibit maximum dexterity at corresponding minimum force and power levels. Arm plus finger-lock combinations on the other hand, give maximum power and force with minimum dexterity.

Table 5. Summary of the Primitive Operations for the Robotic Hand.

1. Movement of the Fingers with Defined Speed
2. Movement of the Fingers With Defined Speed to a Defined Position
3. Grasping With Defined/Specified Force and Torque Limits
4. Movement of the Fingers With Defined Speed to a Defined Position, Combined With the Exertion of a Defined Force
5. Stopping the Fingers at Any Closing Distance. This Could Include Locating the Hand Fingers Concentric With/In Accommodation to - an Object - Using Touch Sensors.
6. Activation of the Sensor System (Tactile and Nontactile)
7. Deactivation of the Sensor System
8. Transfer of the Sensor Data to the Robot Hand Controller

Table 5 summarizes the eight primitive operations required for the robotic hand, which it should be remembered is part of the MMU, arm, hand - EVA Retriever system. Note that

* The first five primitive operations ultimately make closed loop control of (1) Position, (2) Speed, and (3) Grasp force necessary.
* Speed control can be realized via hardware.
* Force and position are controlled by software - which implies the use of dynamics/dynamic models of the robotic hand, arm, MMU system.
* For control of the (grasped/manipulated) object with the robotic hand it is necessary to use both tactile and non-tactile sensors.

Dynamics

Dynamics can be considered to be made up of the two subareas:

kinematics and kinetics. As applied to robotic hands and manipulators, kinematics is concerned with the (a) Direct mapping of joint coordinate motion to link Cartesian coordinate motion, and (b) Inverse mapping of link Cartesian coordinate motion to joint coordinate motion. Here the emphasis is on the hand/manipulator motion without regard to the forces and torques causing the motion. Hand/manipulator kinematics involves the position, velocity, and acceleration aspects of the motion in link Cartesian coordinate and joint coordinate spaces.

Kinetics as applied to robotic hands and manipulators is concerned with the time rate of change of the Cartesian configuration of these robotic systems as a function of the applied joint torques and forces. Here the Newton-Euler or the Lagrangian forms of the equations of motion can be used to define the dynamic response of the hand/manipulator linkage system to input joint forces and torques. Once the equations of motion have been developed, the main problem is to determine, i.e., identify those design parameters and force/torque terms which are important in controlling the Cartesian space motion of the hand/manipulator system. Thus, dynamics is important to the design and the control of the dextrous hand plus arm system. It can lead to better nonlinear and linearized models for simulation tools in task planning and control strategy implementation. This is especially true for the multi-fingered hand plus multi-armed EVA Retriever systems. This section of the report will concentrate on the kinetics aspects of dynamics.

Table 6 summarizes a nonlinear - recurrence formulation of the Newton-Euler equations for a hand manipulator system. This is the Luh, Walker, Paul formulation. Table 7 defines the nomenclature used in Table 6. Details can be found in Luh et al 1983, Lathrop 1985, and Murray and Newman 1986. The corresponding linearized - recursive formulation of the Newton-Euler equations is given in Table 8. The nonlinear Newton-Euler dynamic equations can be used to generate feedforward control torques/forces for the hand plus arm system. Depending upon the computer processing power available, the nonlinear dynamics calculations can be done either offline or online. Multiprocessor implementation of the latter is required for robust control in real-time situations.

The linearized - recursive dynamic equation models can be used for (a) Developing optimal/more-optimal robot hand, etc. designs, (b) Online/real-time identification of unknown object inertia and robot drive train, etc. characteristics, and (c) Implementing observer synthesis giving robust adaptive control.

Implementing the dynamic formalisms given above can lead to direct concern with the following bottlenecks - i.e., operation count and memory impediments to real-time dynamic models. The computational load problem of many advanced control schemes lies in developing an efficient inverse dynamics algorithm. Such algorithms can compute the actuator torques/forces required to produce the desired joint accelerations (a la inverse kinematics) for a given set of link Cartesian displacements. Such algorithms must be evaluated by the control computer at least 100 times per second for the hand/manipulator control scheme to be effective.

Here it is important to exploit the reduction of operation

TABLE 6, SUMMARY OF THE RECURSIVE NEWTON-EULER FORMULATION

Forward Iterations for $i = 1, 2, \ldots, N$.
Initialize: $\omega_0 = \dot{\omega}_0 = 0$ and $\dot{v}_0 = -g$.

$$\omega_i = R_i^T[\omega_{i-1} + \sigma_i z_0 \dot{\theta}_i]$$

$$\dot{\omega}_i = R_i^T\{\dot{\omega}_{i-1} + \sigma_i[z_0\ddot{\theta}_i + \omega_{i-1} \times (z_0\dot{\theta}_i)]\}$$

$$\dot{v}_i = R_i^T[\dot{v}_{i-1} + (1-\sigma_i)[z_0\ddot{d}_i + 2\omega_{i-1} \times (z_0\dot{d}_i)]\}$$
$$+ \dot{\omega}_i \times p_i + \omega_i \times (\omega_i \times p_i)$$

Backward Iterations for $i = N, N-1, \ldots, 1$.
Initialize: $f_{N+1} = n_{N+1} = 0$.

$$\dot{v}_{ci} = \dot{v}_i + \dot{\omega}_i \times r_i + \omega_i \times (\omega_i \times r_i)$$

$$F_i = m_i \dot{v}_{ci}$$

$$N_i = I_{ci}\dot{\omega}_i + \omega_i \times (I_{ci}\omega_i)$$

$$f_i = R_{i+1}[f_{i+1}] + F_i$$

$$n_i = R_{i+1}[n_{i+1}] + p_i \times f_i + N_i + r_i \times F_i$$

$$\tau_i(t) = \sigma_i n_i^T(R_i^T z_0) + (1-\sigma_i) f_i^T(R_i^T z_0)$$

Table 7 Nomenclature

Scalars

- σ_i = Boolean joint type variable: $\sigma_i = 1$ if joint i is revolute; $\sigma_i = 0$ if joint i is prismatic.
- m_i = total mass of link i
- $\tau_i(t)$ = joint torque/force at joint i

3-Vectors (Referenced to the ith Coordinate Frame)

- $\omega_i, \dot{\omega}_i$, and $\dot{v}_i$ = angular velocity, acceleration, and linear acceleration of the ith coordinate frame
- $\dot{v}_{ci}$ = linear acceleration of the center-of-mass of link i
- F_i and N_i = net force and moment exerted on link i
- f_i and n_i = force and moment exerted on link i by link $(i-1)$
- p_i = position of the ith coordinate frame with respect to the $(i-1)$st coordinate frame: $p_i = [a_i, d_i \sin(\alpha_i), d_i \cos(\alpha_i)]^T$
- r_i = position of the center-of-mass of link i: $r_i = [r_{xi}\ r_{yi}\ r_{zi}]^T$
- z_0 = $[0\ 0\ 1]^T$

(3×3) Matrices

- R_i = orthogonal rotation matrix which transforms a vector in the $(i+1)$st coordinate frame to a coordinate frame which is parallel to the ith coordinate frame:

$$R_i = \begin{bmatrix} \cos(\theta_i) & -\cos(\alpha_i)\sin(\theta_i) & \sin(\alpha_i)\sin(\theta_i) \\ \sin(\theta_i) & \cos(\alpha_i)\cos(\theta_i) & -\sin(\alpha_i)\cos(\theta_i) \\ 0 & \sin(\alpha_i) & \cos(\alpha_i) \end{bmatrix}$$

for $i = 1, 2, \ldots, N$, where $R_{N+1} = I$.

- I_{ci} = classical inertia tensor [13] of link i about a coordinate frame parallel to the ith coordinate frame and translated to the center-of-mass of link i; with principal inertias I_{xxi}, I_{yyi}, and I_{zzi}; and cross-inertias I_{xyi}, I_{xzi}, and I_{yzi}.

TABLE 8 SUMMARY OF THE RECURSIVE LINEARIZED NEWTON-EULER FORMULATION

Forward Iterations for $i = 1, 2, \ldots, N$.
Initialize: $\delta\omega_0 = \delta\dot{\omega}_0 = \delta\dot{v}_0 = 0$.

$$\delta\omega_i = R_i^T[\delta\omega_{i-1} + \sigma_i[z_0\delta\dot{\theta}_i - Q(\omega_{i-1} + z_0\dot{\theta}_i)\delta\theta_i]]$$

$$\delta\dot{\omega}_i = R_i^T[\delta\dot{\omega}_{i-1} - \sigma_i[z_0\delta\ddot{\theta}_i + \delta\omega_{i-1} \times (z_0\dot{\theta}_i)$$
$$+ \omega_{i-1} \times (z_0\dot{\delta\theta}_i)]$$
$$- \sigma_i Q\{\omega_{i-1} + z_0\ddot{\theta}_i + \omega_{i-1} \times (z_0\dot{\theta}_i)\}\delta\theta_i]$$

$$\delta\dot{v}_i = R_i^T[\delta\dot{v}_{i-1} - \sigma_i Q\dot{v}_{i-1}\delta\theta_i + (1-\sigma_i)[z_0\delta\ddot{d}_i$$
$$+ 2\delta\omega_{i-1} \times (z_0\dot{d}_i) + 2\omega_{i-1} \times (z_0\delta\dot{d}_i)]]$$
$$+ \delta\dot{\omega}_i \times p_i + \delta\omega_i \times (\omega_i \times p_i) + \omega_i \times (\delta\omega_i \times p_i)$$
$$+ (1-\sigma_i)(\dot{\omega}_i \times p_{di} + \omega_i \times (\omega_i \times p_{di}))\delta\theta_i$$

Backward Iterations for $i = N, N-1, \ldots, 1$.
Initialize: $\delta f_{N+1} = \delta n_{N+1} = 0$.

$$\delta\dot{v}_{ci} = \delta\dot{v}_i + \delta\dot{\omega}_i \times r_i + \dot{\omega}_i \times (\omega_i \times r_i) + \omega_i \times (\delta\omega_i \times r_i)$$

$$\delta F_i = m_i \delta\dot{v}_{ci}$$

$$\delta N_i = I_{ci}\delta\dot{\omega}_i + \delta\omega_i \times (I_{ci}\omega_i) + \omega_i \times (I_{ci}\delta\omega_i)$$

$$\delta f_i = R_{i+1}[\delta f_{i+1}] + \delta F_i + QR_{i+1}[f_{i+1}]\delta\theta_{i+1}$$

$$\delta n_i = R_{i+1}[\delta n_{i+1}] + \delta N_i + p_i \times \delta f_i + r_i \times \delta F_i$$
$$+ (1-\sigma_i)(p_{di} \times f_i)\delta d_i + \sigma_{i+1} QR_{i+1}[n_{i+1}]\delta\theta_{i+1}$$

$$\delta\tau_i(t) = \sigma_i[\delta n_i^T(R_i^T z_0) - n_i^T(R_i^T Q z_0)\delta\theta_i]$$
$$+ (1-\sigma_i)[\delta f_i^T(R_i^T z_0)]$$

NOTE THAT THIS LINEARIZED DYNAMIC ROBOT MODEL CAN BE USED FOR REAL TIME DIGITAL SIMULATION - OBSERVER SYNTHESIS TO GENERATE A FEEDBACK SIGNAL THAT REGULATES THE PERTURBATION POSITIONS AND VELOCITIES TO ZERO. IT CAN ALSO BE USED FOR IDENTIFICATION OF THE MANIPULATOR. ETC. PARAMETERS VIA THE LINEARIZED DYNAMIC ROBOT MODEL

TABLE 9 SUMMARY COMPARISON OF DYNAMICS FORMULATION (ADAPTED FROM HOLLERBACH 1980)

Method	Comparison of Time Dependences		Comparison for $n = 6$	
	Mults.	Adds.	Mults.	Adds.
Uicker/Kahn (original Lagrangian)	$32\frac{1}{3}n^4 + 86\frac{1}{3}n^3 + 171\frac{1}{3}n^2 + 53\frac{1}{3}n - 128$	$25n^4 + 66\frac{1}{3}n^3 + 129\frac{1}{3}n^2 + 42\frac{1}{3}n - 96$	66,271	51,348
Waters (partially recursive)	$-106\frac{1}{3}n^3 + 620\frac{1}{3}n - 512$	$82n^3 + 514n - 384$	7,051	5,652
Hollerbach (4 × 4 Lagrangian)	$830n - 592$	$675n - 464$	4,388	3,586
Hollerbach (3 × 3 Lagrangian)	$412n - 277$	$320n - 201$	2,195	1,719
Luh, Walker, Paul (Newton-Euler)	$150n - 48$	$131n - 48$	852	738
Horn, Raibert (table lookup)	$2n^2 + n^2$	$n^2 + n^2 + 2n$	468	264
Luh, Lin (scheduled parallel Newton-Euler)	$57n - 18$ (estimated; see text)	$50n - 18$ (estimated)	323 (estimated)	280 (estimated)
Lathrop (linear parallel Newton-Euler)	$2n + 3$	$6n + 7$	15	43
Lathrop (logarithmic parallel Newton-Euler)	$2\lfloor\log_2(n+1)\rfloor + 5$	$6\lfloor\log_2(n+1)\rfloor + 10$	11	28
Lathrop (systolic pipeline)	1 (successive; see text)	3 (successive)	1 (successive)	3 (successive)

Note: This table reflects the algorithmically indicated cost for the fully general six-link rotary manipulator, as in Hollerbach (1980). By consideration of special cases, introduction of coordinates or workspace assumptions, or tailoring of the computations, additional reductions are possible.

Tables I & II from J. M. Hollerbach, "A Recursive Formulation of Manipulator Dynamics," *IEEE Trans. Systems, Man & Cybernetics*, SMC-10, No. 11, pp. 730–736, have been adapted to include new algorithms since 1980.

count/time of calculation as a function of the number of degrees of the hand/manipulator robotic system. Such reductions arise from the geometry and hierarchical design structure (as, e.g., open tree and constrained loop - connectedness partitioning) of the smart hand plus arm robotic system.

Horak 1984 has considered the exploitation via the arm, wrist/hand partitioning of the relatively simple geometry of arm - manipulator systems having 4 to 7 links. Lathrop 1985 has treated the exploitation of parallelism in manipulator dynamics. Table 9 summarizes his results. It also shows the evolution toward reduced operation (addition, multiplication) and machine cycle time. Here as with the forward/backward recurrence schemes in Tables 6, 8, linear operation counts are possible. However by exploiting the space-time parallelism (a) Order log(NDOF) operations and (b) Systolic 3 vector adds plus 1 vector multiply implementations are also possible. Lathrop's work indicates that such efficiencies obtain for both the nonlinear and the linearized dynamic equation forms.

As far as the control of object motion for manipulators is concerned, Kerr and Roth 1986 have analyzed the inverse dynamics problem for multi-fingered hands. They give a general formulation to evaluate the finger, etc. torques/forces required for the redundant object plus multi-finger linkage system. The author has considered the degree of freedom reductions possible if position and velocity/slip sensors are used with each finger. Details will be available in an extended version of this report.

Sensors and Sensing Systems

Sensors and sensing systems are the beginning elements or components in perception. They are of prime importance in performing the observation function - for the intelligent, controlled interaction of a robotic system with its environment. The primary sensors/sensing systems considered for use in general space robotics can be summarized as (1) Vision, (2) Tactile, (3) Proximity, (4) Force, (5) Position, and (6) Velocity.

The smart, dextrous hand plus arm plus MMU - EVA Retriever system needs a variety of sensors. Multi-sensing is necessary in order for the Retriever robotic system to adapt to disturbances and unpredictable changes in (a) Its EVA - space environment, and (b) Itself in interacting with that environment. Here the Retriever robotic system will use (1) External sensors as, e.g., visual, infrared - light, solid-state CCD cameras and touch sensors; and (2) Internal sensors such as optical, magnetic-joint encoders, and Hall-effect based, joint torque/force sensors. In this vein, Table 10 summarizes the sensor environment of the robotic hand, etc system. Figure 7 gives a corresponding representation of a robotic manipulator/hand system which is provided with the senses of feeling (internal sensing) and sight (external sensing). Figure 7a also gives an operational diagram for the (1) Scene recognition, (2) Signal/image processing - programs, and (3) Grasp program for the manipulator hand. Figure 7b depicts the teaching-phase use of the sensed information.

In the smart, dextrous robotic hand plus arm system, sensors can measure triaxial force at the finger/hand contact points with an ob-

TABLE 10. SUMMARY OF THE SENSOR ENVIRONMENT OF THE ROBOTIC HAND.

1. Tactile Sensors As Per:
   - Force-Sensing Wrist
   - Grasp Force Sensor
   - Touch Sensors
2. Nontactile Sensors As Per:
   - Object Approach Sensors (Farfield, Midfield Vision)
   - Object 'Capture' Sensors (Nearfield Position Recognition Sensors, As for Example, Infrared for Determining Object Position Between the Fingers)

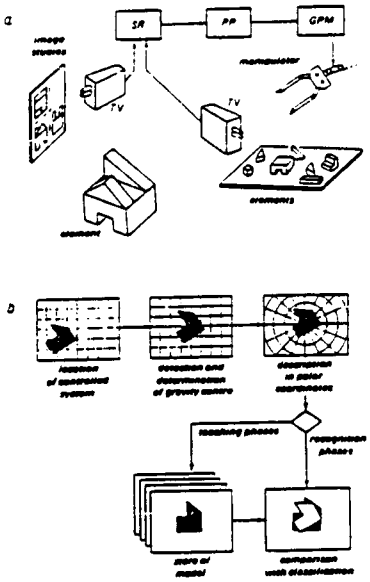


FIGURE 7. REPRESENTATION OF ROBOTIC MANIPULATOR/HAND PROVIDED WITH SENSES OF FEELING (INTERNAL SENSING) AND SIGHT (EXTERNAL SENSING). FIGURE a GIVES THE OPERATIONAL DIAGRAM WITH (1) SR= SCENSE RECOGNITION, (2) PP= PROCESSING PROGRAMS, AND (3) GPM= GROUP PROGRAM FOR THE MANIPULATOR HAND; FIGURE b DEPICTS THE TEACHING PROCESS.

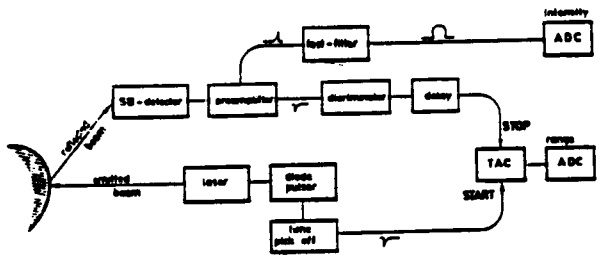


FIGURE 8. REPRESENTATION BLOCK DIAGRAM FOR A TIME-OF-LIGHT LASER SCANNER

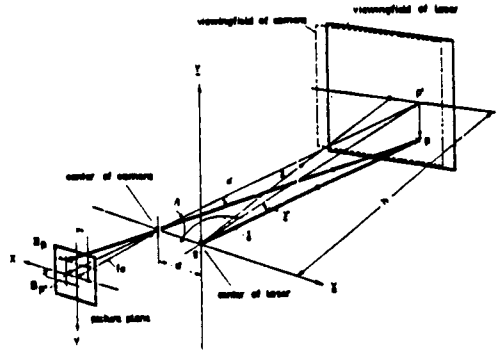


FIGURE 9. BASIC GEOMETRY OF A TRIANGULATION TYPE LASER SCANNER

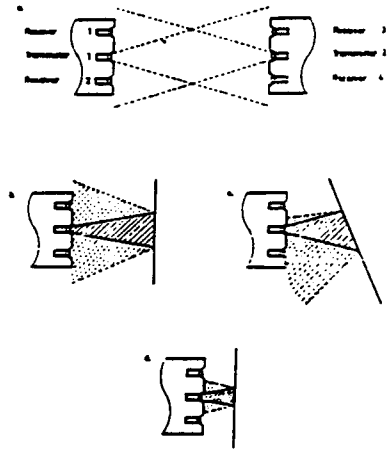


FIGURE 10. DIPICTION OF THE PRINCIPLE OF INFRARED DETECTION OF AN OBJECT BETWEEN FINGERS
(A) CONFIGURATION FOR INFRARED TRANSMITTER AND RECEIVER
(B and C) REFLECTION OF SIGNALS BY PARALLEL AND NON PARALLEL OBJECT SURFACES, RESPECTIVELY.
(D) SENSOR CONFIGURATION FOR OBJECT APPROACH SITUATION

ject [Stocic et al 1986]. The object will be grasped/manipulated - as, e.g., while either in motion during a carry/transport operation or in contact with the working environment, as in an assembly task. In this context, sensing can allow the robotic system to identify and then compensate for - uncertainties in the object inertia related - load and torque characteristics [Neuman, Khosla 1986, Seraji 1987, Stephenko 1987].

For sensor-based, effective task planning and control operations, the following dynamic world data are used for trajectory calculations: (1) Position and orientation of the retrieved object, (2) Distances, (3) Diameters, (4) Geometry of the retrieved object, (5) Path of the object (position, velocity), and (6) Contours. Vision (visual light and infrared) sensors provide such data for the EVA Retriever robotic system. Figure 8 gives a block diagram for a time-of-light laser scanner. Figure 9 depicts the basic geometry for a triangulation type laser scanner. These sensor systems can be used in mid-field to farfield vision systems to obtain rest of the dynamic world data discussed above. Up close - that is, in the near field of the robotic hand between the finger - arrays of infrared sensors can be used to sense object approach/proximity and orientation. Figure 10 depicts this use of infrared receiver, transmitter arrays.

Smart hand tactile sensing is required when the intelligent EVA robotic retriever system must: (a) Verify capture and stable grasping of an object, and (b) Perform delicate manipulate/assembly operations. It should be noted that the tactile sensors can make many manipulate/assembly etc. tasks simple which are very difficult to accomplish visually.

Tactile-sensing reduces to the following three fundamental sensing operations:

(1) Joint forces sensing - ie, sensing the forces applied to the joint of the robotic hand, wrist, and arm manipulator).
(2) Touch-sensing of the pressures applied at important contact points on the finger/hand surface.
(3) Slip-sensing of movement/incipient movement of the object while it is being grasped.

Joint forces are typically sensed using arrays of strain gages (strain-gage wrist force sensor) or Hall effect sensors,(fingers, etc. of the UTAH/MIT dextrous hand). Vibration or relative displacement sensors are used to detect object slip [Staugard 1987].

The sense of touch is used to generate the finer detail data required to describe the interaction between the robotic hand and the object or workpiece. This is because most of the manipulative/assembly action occurs at the contact interface between the hand and the object. A robot touch sensor should duplicate the object recognition, and position and orientation determination functions of the human skin/nerve ending system. Usually temperature sensing per se is not required. Table 11 summarizes the design specifications/characteristics of a representative tactile sensor. Figures 11 and 12 give schematic representations of the Hillis conductive rubber resistance change, and the Raibert and Tanner VLSI tactile/touch sensor array designs, respectively. Acoustic pulse through tuned elastomer array designs from Bonne-

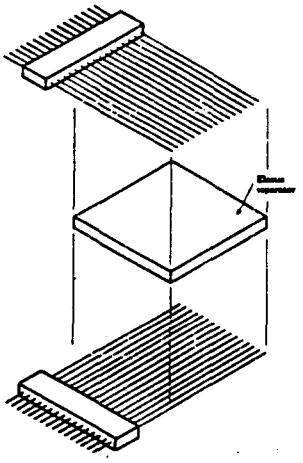


FIGURE 11. KILLIS DESIGN TACTILE SENSOR. THIS SENSOR CONSISTS OF A MONOLITHIC ARRAY OF 256 INDIVIDUAL SENSORY ELEMENTS. TYPICAL DESIGNS HAVE TWO SHEETS OF WIRES RUNNING PERPENDICULAR TO EACH OTHER SEPARATED BY A THIN ELASTIC MEDIUM. HERE THE INTERSECTION POINTS OF THE WIRES FORM THE INDIVIDUAL SENSORY ELEMENTS FOR PRESSURE MEASUREMENT.

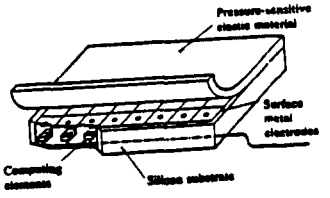


FIGURE 12. PHYSICAL STRUCTURE OF THE RAIBERT AND TANNER VLSI TACTILE ARRAY SENSOR DESIGN. HERE A 1-mm THICK SHEET OF PRESSURE-SENSITIVE CONDUCTIVE PLASTIC (DYNACION ®) IS PLACED IN CONTACT WITH A VLSI WAFER (AN NMOS INTEGRATED CIRCUIT. THE WAFER IS COMPRISED OF A TWO-DIMENSIONAL ARRAY OF CELLS, EACH OF DIMENSION 1.6 x 0.9 mm.

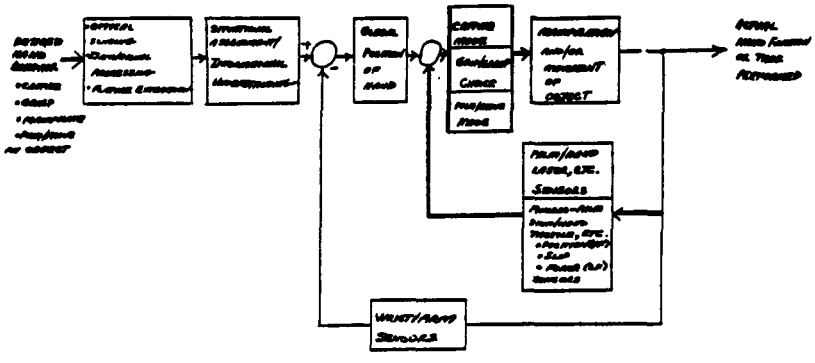


FIGURE 13. SCHEMATIC REPRESENTATION OF SIMPLE HAND MANIPULATION PROCESS AS A FEEDBACK OPERATION

TABLE 11. SUMMARY OF THE CHARACTERISTICS/DESIGN SPECIFICATIONS OF A REPRESENTATIVE TACTILE TRANSDUCER.

1. An Array Consisting of at Least 10 X 10 Force-Sensing Elements on a 1 sq-in Surface. This Corresponds to a Human Fingertip.
2. Each Element Should Have a Response Time of 1 to 10 msec. This should Be Closer to 1 msec. (Corresponds to 500 Hz Via Nyquist Sampling Criteria).
3. Threshold Sensitivity for the Elements Should Be Alpha X 1 g , With the Upper Limit of the Force Range at Alpha X 1000 g . Here Alpha Is a Factor Appropriate to the Space Environment (0 << Alpha < 1).
4. The Elements Do Not Need to Be Linear. However, They Must Have Low Hysteresis.
5. The Skinlike Material Must Be Robust. That Is, It Must Stand Up Well to the Harsh Space Environment and Be 'Modularly' Replaceable.

TABLE 12. FACTORS TO BE CONSIDERED BY A LANGUAGE FOR A VISION AND MULTISENSOR SYSTEM.

1. Image/Signal Preprocessing (Including Restoration, Enhancement, Etc.)
   - Instructions for Component 'Visual' and Multi-sensor Systems
   - Processing of All 'Pixel' or Array Element Level Information
   - Parallel Local Processing Operation
   - Efficient Storing of Multidimensional (2D,3D,Etc.) Structured Data
2. Image/Signal Extraction (Feature Extraction, Object Recognition Including MultiSensor Fusion)
   - Arithmetic, Delay Operations
   - Statistical Operations - Fusion of MultiSensor Information
   - Instructions for the 'Teach-In'/Learning Mode
   - Instructions for Feature Definition
   - Quantitative Feature Extraction
3. Image Analysis (Description)
   - Definition of Data Structures (Lists, Graphs, Etc.)
   - Definition of Object Value Relations ('Above', 'Below', 'Left', 'Right', Etc.)
   - Composition (Decomposition) of Relations
   - Knowledge Acquirement (Knowledge Accretion)
   - Model Generation (Via Geometric, Symbolic, CAD, Etc. Representations)
4. Image/Signal Context ('Scene' Analysis, Including Situation Assessment/Informational Understanding)
   - Analysis Based on Stored Knowledge
   - Inference Rules
   - Reasoning and Queries
   - Knowledge Management

Note that Points 3 and 4 Motivate the Construction of Symbolic, Rule-Based Languages Derived from the Results of Artificial Intelligence Research.

vile Scientific [Astle 1987] also look promising.

In considering the use of sensors in effective, task planning and control strategies for an intelligent hand, etc. Retriever robotic system, the computer language used must be given great consideration. Table 12 gives the factors to be considered by a computer programming language for a vision and multisensor system. Here the four main areas of (1) Image/Signal Processing, (2) Image/Signal Extraction, (3) Image/Analysis, and (4) Image/Signal Context (Scene Analysis) are considered. Requirements in these areas drive the design specification features which must be offered in a multi sensor-based, intelligent control system.

Retriever Robotic Control

Intelligent, sensor-based control of the smart dexterous hand plus arm plus MMU components of the Retriever robotic system is of primary importance to its successful EVA function. Here the arm plus hand/manipulator tasks of the Retriever can be specified in Cartesian coordinates. As discussed in the previous report section, sensors perform the observation function. The measured internal variable for feedback control purposes are joint displacements and velocities. Thus, the goal or command state variables and these measured output variable quantities are in different coordinates. This means that control of the position and orientation of the arm and hand/manipulator system by actuators at the joints requires knowledge of the transformation between the Cartesian hand and joint space respresentations.

Figure 13 gives a schematic representation of the smart hand prehension process as a feedback operation. In the system depicted desired hand behavior task commands are used in combination with global sensing to position the arm/hand system. At this level the hand mode is chosen and used to bring about grasping/low-level and /or movement of the object. Sensor information is continually fed back to control finger and hand object position, velocity and contact force.

Deep Knowledge Control

In developing grasp and manipulate - "trajectory" control of the path of a retrieved object, or an object/robotic hand system in contact with the environment, the dynamic modeling and the sensing functions discussed previously can be exploited. The hand/manipulator dynamics can be defined in joint space coordinates as

$$H\ddot{q} + h = T \qquad (1)$$

in which, q = Vector of joint displacements (θ, d) as discussed previously.

T = Vector of joint torques.

$H(q)$ = Manipulator inertia matrix.

$h(q, \dot{q}, t)$ = Nonlinear term containing the centrifugal, Coriolis, friction, and gravitational forces/torques.

As discussed previously, sophisticated on-line computational strategies are required to compute the arm plus hand/manipulator dynamics for the EVA Retriever robotic system. In today's arm plus

TABLE 13. SUMMARY OF THE NECESSARY FEATURES OF A MULTIPROCESSOR-BASED ROBOT CONTROLLER FOR THE EVA RETRIEVER.

1. Different Control Levels (Including Task Accomplishment, Opportunistic Scheduling)
2. Distribution of the Control Hierarchy for the Robot Among High-Speed, 32 Bit Microprocessors
3. Distributed System Operation (Necessary for Parallel Task Processing on Different Control Levels)
4. Modular Symmetric Hardware Configuration
5. Multiple Bus System
6. Expandability With No Need for Hardware or Software Reconfiguration
7. Special-Purpose Modules (Sensor Input/Output, Arithmetic, FFT, Interpolators, Kalman Filters for Observation and Control, Etc.)
8. Parallel Task Decompositions (A La Brooks, Lozano-Perez) for Real-Time/Real-World Problem Situations
9. Handling of a World Model in Global Memory

TABLE 14. BASIC COMPUTER ARCHITECTURE REQUIREMENTS - UNDER WHICH DEVELOPMENT AND IMPLEMENTATION OF THE ROBOT CONTROL SYSTEM SHOULD BE PERFORMED.

1. Clear, Well-Defined Separation of All Control Levels
2. Modular Architecture Built-Up on the Basis of a Symmetrically Distributed System
3. Standardized Interfaces Between the Control Modules (Hardware and Software)
4. Transparency to the User at All Control Levels
5. Defined Format and Standardized Protocols
6. Multibus System to Ensure Parallel Communication
7. Hierarchically-Based, 'Decentralized' System Control
8. Dynamic Reconfiguration for a (High Signal Communication to Processing Power - A La Hypercube Type) Polyprocessor Environment

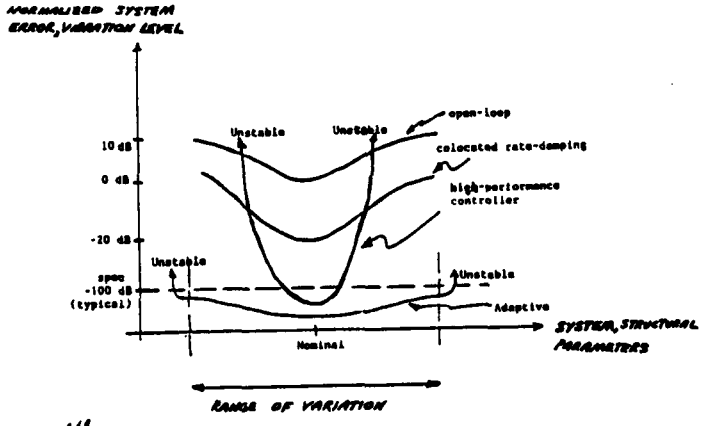


FIGURE 14. CLOSED-LOOP PERFORMANCE VERSUS VARIATION IN SYSTEM, STRUCTURAL PARAMETERS

FIGURE 15. (a) REPRESENTATION OF THE TRADITIONAL DECOMPOSITION OF A MOBILE ROBOT CONTROL SYSTEM INTO FUNCTIONAL MODULES ('STAFF' ORGANIZATION)

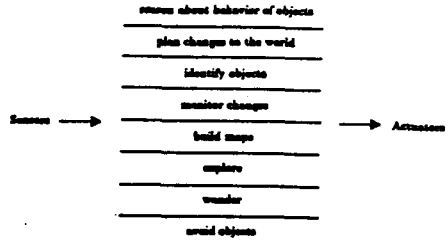


FIGURE 15. (b) REPRESENTATION OF A ROBOT CONTROL SYSTEM BASED ON TASK-ACHIEVING BEHAVIORS ('PROJECT' ORGANIZATION)

hand, etc. systems for the EVA Retriever the trend is toward mechanically "cleaner", more positive designs with fewer gear, etc. drive train components. This means that actuators, as, e.g., high torque motors are coupled directly or as closely as possible - to each joint in direct drive robotic system designs. Such designs have high mechanical stiffness, little or no backlash, and low friction. However, they also can have higher sensitivity to external disturbances, full link-to-link coupling, and nonlinear dynamics effects which are made worse by the high speeds which such systems can attain. These factors can increase the need for on-line computational dynamics and control.

Controlling robotic system dynamics of the form of Eq. (1) usually involves one of the three approaches:

(1) Individual joint proportional, integral, derivative (PID) control - which (a) Ignores link dynamics, and (b) Uses local decoupled PID's at each joint (ignores link coupling).

(2) Computed torque control - in which control torque T is defined using a structure identical to Eq. (1). That is,

$$T = H\,u + h \tag{2}$$

Here the problem reduces to that of controlling the system

$$\ddot{q} = u \tag{3}$$

Such control can be done using decoupled PID as per:

$$u = \ddot{q}_{Des_j} - KD_j\,\dot{\tilde{q}}_j - KP_j\,\tilde{q}_j - KI_j\int_0^t \tilde{q}_j\,dt \tag{4}$$

in which, $\tilde{q}_j = q_j - q_{Des_j}$.

KP_j, KI_j, KD_j = PID coefficients which are positive and sufficiently large.

(3) Robust Controller Design - as, e.g., using pole-placement modal control and gain scheduling in combined feedforward plus feedback control strategies [Norcross et al 1986, Seraji 1987, Stephenko 1987]. Here the basic issue is how to minimize performance sensitivity to model model uncertainty, as, e.g., parameter uncertainty or inaccuracies in manipulator and object mass properties, torque constants of the actuators, friction of the drive train, unknown loads, etc. Model uncertainty can also encompass high-frequency unmodeled dynamics - such as structural resonant modes, sampling rates, and neglected time delays.

Figure 14 illustrates the idea of controlled system robustness. It plots closed-loop performance versus the variation in system parameters. This is done for different types of controlled dynamic system - ranging from open-loop (no feedback control) to adaptive (model reference ,parameter estimation, etc. types), closed-loop control.

Relevant control of the EVA Retriever robotic system it should be noted that there is an important tradeoff between (1) Speed, (2) Cost, and (3) Generality of application. This motivates the selecton of a two-level control structure as per: (1) Linear low level

(inner loop) feedback control law, and (2) Generally nonlinear upper level (outer loop) feedforward control law. The outer loop optimizes the linear feedback control law parameters. This is done based upon the nonlinear equations of motion for the robotic system.

Hierarchical Strategy and Intelligent Control

In implementing control of the hand plus arm plus MMU comprising the EVA Retriever robotic system, it is important to consider the strategy/control architecture employed. Figure 15a gives a representation of the traditional decomposition of a mobile robot control system into functional modules. This "staff" organization of the control system is inadequate for the real-time implementation on a multiprocessor of control for the EVA Retriever. Figure 15b is a representation of a multilayered robot control system based on task-achieving behaviors. This "project" organization of the control system is made up of hierarchically parallel task layers. It lends itself to implementation on a multiprocessor based controller and hence shows great promise for real-time control of the complete EVA Retriever system [Brooks 1986].

Basic intelligent control algorithms can be implemented using a parallel microprocessor (i.e., multiprocessor based) structure. Here such rule-based algorithms for MIMO (multi-input, multi-output) systems have the property of decomposing or decoupling the multivariable control input into a set of single-input, multi-output systems. Thus, this allows for a parallel, pipelined (i.e., systolic) structure using a common rule base (as, e.g., with cloned CLIPS, etc.) that is easily implementable on an array of processors.

Any discussion of smart robotic hand control would be incomplete without mentioning the possibility of using neural nets [Lippmann 1987] or generalized learning algorithms [Miller et al 1987]. These can be used for multi-sensor pattern recognition functions and for the control of robotic hand/manipulator systems. Here the pattern recognition and control-learning schemes can be based solely upon observations of the input-output relationship for the system being controlled. Research work with Hopfield type neural nets [Freeman 1987] used in solving optimization problems, indicates that control objectives can be included. The author suggests the formulation of a generalized Hopfield neural net structure. This would include optimal robotic controls imbedding - as a Hamiltonian/Lagrangian form, augmented with state vector and control vector constraints.

Multiprocessor-Based Robot Control

The multiprocessor implementation of real-time intelligent control of the hand plus arm plus MMU comprising the EVA Retriever is a function of such issues as:

* Computational load balance
* Tradeoffs between (a) Signal communication time, and (b) Processing time
* Programming techniques - as a function of (a) Hardware/software architecture, and (b) Language environment
* Effective raw processor speed, quantization, and memory utilization
* "Impedance Match" with the problem formulation and algorithmic structure and complexity.

Table 13 summarizes the necessary features of a multiprocessor-based robotic controller for the EVA retriever. The basic computer architecture requirements - under which development and implementation of the robotic control system should be performed are given in Table 14. Both tables emphasize features/requirements relating to doing parallel task processing on different control levels.

Conclusions

An investigation has been made of the basic issues and concepts relating to dynamics, sensing, and intelligent control for the dextrous hand plus arm plus MMU system. This system defines the space-going components of the EVA Retriever robotic system. Specific attention has also been given to artificial intelligence, rule-based Expert Systems and the new area of Neural Nets. Based upon the research conducted, the author considers the following of prime importance: (1) Integrated dextrous hand plus articulated arm, mechanical and electrical drive system design, (2) Multisensor capabilities for position, velocity, force/torque, and slip/grasp stability, (3) "Standard" control and neural net observer synthesis for control purposes, and (4) Multiprocessor-based implementation of real-time, on-line dynamics, robust/adaptive control and hierarchally-layered (Brooks architecture) - task planning. Specific recommendations are summarized in the next report section in terms of a research and development program plan.

Recommendations

This section of the report presents recommendations for the smart robotic hands plus arms plus MMU - EVA Retriever system. These recommendations are presented in the form of a research and development program plan. The R & D program plan gives activities that can be pursued on a one, two, and three year basis. A more complete version, broken down into specific research and development plans will be available in an extended version of this report.

(1) Implementation of basic primitive operations for the robotic hands: (a) open-loop with force limit, and (b) Sensor-hand -- low level (basic movement) and high level (position, force contact, and slip/grasp stability) control.

(2) Implementation of task planner for six grasp and six basic manipulate robotic hand functions.

(3) Neural Nets developmental work: (a) Testbed w.r.t. robotic hands applications (pattern recognition for rule-based systems, control primitives, basis for "common sense" reasoning, use in fuzzy/noisy situations), (b) Use with A.D. Little/Marcus hand - master system, (c) Developmental operating system,(d) Knowledge representation significance versus sentence-like representations of knowledge, (e) Explanation of Neural Net reasoning and knowledge representation utility.

(4) Development of multiprocessor, systolic - type transputer versions of CLIPS, Common LISP: (a) "Cloned" CLIPS, etc. architecture with "Guarded Horn Clause" resolution of

contention for system resources, (b) Interfaced with multiprocessor versions of C and FORTRAN.

(5) Development of integrated AI/ES & NN plus digital signal processing system for using multisensor information.

(6) Development of distributed operating system which is necessary for/mandated by - the need to do parallel task processing on different hierarchical control/task levels.

(7) Development of ongoing testing and performance evaluation: (a) Prototyping tools as per (6) above, and (b) hardware, software AI/ES & NN controller testbed (hands etc.) for EVA Retriever (and Space Station intelligent control applications.

REFERENCES

1. "Extravenicular Activity (EVA) Retriever Program Plan," NASA/ Johnson Space Center, Houston, Texas, Report No. JSC-22144, May 1987.
2. L. Astle, Presentation to CTSD - NASA/JSC on Bonneville Scientific - Acoustic Tactile Sensor Array and Related Micromechanical and VLSI Capabilities, July 1987.
3. J. Freeman, "Neural Nets Seminar," NASA/JSC, July-August 1987.
4. D. Hillis, "Active Touch Sensing," International Journal of Robotics Research, Vol. 1, No. 2, 1982, pp 33-44.
5. D.T. Horak, "A Simplified Modeling and Computational Scheme for Manipulator Dynamics," Journal of Dynamic Systems, Measurement, and Control, Vol. 106, No. 4, December 1984, pp 350-353.
6. J. Kerr, B. Roth, "Analysis of Multifingered Hands," International Journal of Robotics Research, Vol. 4, No. 4, Winter 1986, pp 3-17.
7. R.H. Lathrop, "Parallelism in Manipulator Dynamics," International Journal of Robotics Research, Vol. 4, No. 2, Summer 1985, pp 80-102.
8. R.P. Lippmann, "An Introduction to Computing With Neural Nets," IEEE Acoustics, Speech, and Signal Processing (ASSP) Magazine, April 1987, pp 4-22.
9. J.Y.S. Luh, M.W. Walker, R.P. Paul, "On-Line Computational Scheme for Mechanical Manipulators," Journal of Dynamic Systems, Measurement, and Control, Vol. 102, 1980.
10. W.T. Miller, et al, "Application of a General Learning Algorithm to the Control of Robotic Manipulators," International Journal of Robotics Research, Vol. 6, No. 2, Summer 1987, pp 84-98.
11. J.J. Murray, C.P. Neuman, "Linearization and Sensitivity Models of the Newton-Euler Dynamic Robot Model," Journal of Dynamic Systems, Measurement, and Control, Vol. 108, No. 3, September 1986, pp 272-276.
12. C.P. Neuman, K.P. Khoshla, "Identification of Robot Dynamics: An Adaptation of Recursive Estimation," In K.S. Narendra (Editor), Adaptive and Learning Systems: Theory and Applications, Plenum, New York, 1986, pp 175-194.
13. R.J. Norcross, et al, "Pole-Placement Methods for Multivariable Control of Robotic Manipulators," Journal of Dynamic Systems, Measurement, and Control, Vol. 108, No. 4, December 1986, pp 340-345.
14. M.H. Raibert, J.E. Tanner, "Design and Implementation of a VLSI Tactile Sensing Computer," International Journal of Robotics Research, Vol. 1, No. 3, 1982, pp 3-18.
15. H. Seraji, "An Approach to Multivariable Control of Manipulators," Journal of Dynamic Systems, Measurement, and Control, Vol. 109, No. 2, June 1987, pp 146-154.
16. A.C. Staugaard, Robotics and AI: An Introduction to Applied Machine Intelligence, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
17. Y. Stephenko, "Modal Control of Fast Large-Scale Robot Motions," Journal of Dynamic Systems, Measurement, and Control, Vol. 109, No. 2, June 1987, pp 80-87.
18. D. Stocic, M. Vukubratovic, D. Hristic, "Implementation of Force Feedback in Manipulation Robots," International Journal of Robotics Research, Vol. 5, No. 1, Spring 1986, pp 66-76.

SPECTROSCOPIC INVESTIGATIONS AND MEASUREMENTS OF CERTAIN ARC JET PARAMETERS

Final Report
NASA/ASEE Summer Faculty Fellowship Program--1987
Johnson Space Center

Prepared By:	Vedula S. Murty
Academic Rank:	Assistant Professor
University and Department:	Texas Southern University Department of Computing and Applied Sciences Houston, Texas 77004
NASA/JSC	
Directorate:	Engineering
Division:	Structure and Mechanics
Branch:	Thermal
JSC Colleague:	John Grimaud
Date:	August 14, 1987
Contract Number:	NGT-44-001-800

ABSTRACT

The NASA/JSC Atmospheric Re-entry Materials and Structures Evaluation Facility (ARMSEF) has been intensively and extensively been involved in ground testing of spacecraft materials and thermal protection systems (TPS) in simulated re-entry conditions. Ground experiments on surface catalytic efficiency of such TPS requires a knowledge of the flow system in the arc jet. In the work described in this report, spectroscopic diagnostic techniques are used to determine the free stream constituents. Specifically, the emission spectrum of the free stream constituents was obtained and the species therein identified.

A laser system has been added on, which will give the added capability of studying the arc jet flow using Laser Raman Spectroscopy (LRS). The LRS technique will complement information obtained from the emission spectra.

A short list of further work that can be done in the area of spectroscopic investigations on the arc jet is mentioned.

INTRODUCTION

The arc jet facility, located at NASA/JSC in Building 222, has played a varied role in simulating re-entry conditions for the purposes of testing spacecraft materials and models in a simulated environment on the ground. The facility consists of two test legs designated TP1 and TP2 arc tunnels. Each test leg consists of:

a. an arc heater
b. test nozzles
c. test chamber with model insertion system
d. a diffuser and a heat exchanger
e. a four stage steam ejector vacuum pumping system

The arc heater is a device for converting electrical energy into thermal energy. A continuous high voltage direct current is established between a cathode and an anode at either end of a segmented column. Gases are then injected into the column and heated by the arc. The heated high energy gas is then expanded through a nozzle to produce a very high supersonic or hypersonic gas stream. The gases used are typically a mixture of nitrogen and oxygen, although it is possible to use other gases.

As explained above, TPS are among one of the many test objects studied at this facility. In order to fully understand the catalytic efficiencies of these TPS (1-3), it is necessary to have a complete knowledge of the flow chemistry, species concentrations, shock temperatures, etc.

Although some of these quantities can be obtained by the strict use of the laws of fluid dynamics, it has been amply demonstrated that spectroscopic techniques (4-5) can also be used to determine some of these quantities. The spectroscopic techniques are non-intrusive and provide a means of making in-situ measurements. At present spectroscopic methods are being used to study only the TP2 component of the facility.

SPECTROSCOPIC SET-UP DESCRIPTION AND EXPERIMENTAL PROCEDURE

A block diagram of the spectroscopic facility is given in Fig. 1. As can be seen from the figure the radiation from the arc jet flow exits the tunnel through a window and is condensed on to mirror M1, which reflects the beam on to a second lens L2, which in turn condenses the light beam on to the slit entrance of the spectrograph. The spectrograph itself is a Spex Triple-Mate spectrograph in which the resolution can be varied to three different settings. The detection is accomplished by a 1024 element linear diode array. The output is recorded on a magnetic disc using an EG&G/PARC optical multichannel analyzer (OMA), which also performs other needed calculations.

In the preliminary set of experiments, a 600 line/mm grating was used which yielded a pixel resolution of 0.069 nm. The first step in the experimental procedure was to calibrate the instruments. The calibration was accomplished in two steps. First the relative sensitivity of the instrument (intensity calibration) was carried out using a standardized carbon-filament, which was operating at a constant current of 38.0A. The wavelength calibration was accomplished using standard Oriel pen lamps. In the wavelength calibration procedure, usually three or four standard lines are chosen and a linear relationship between their cursor position and the wavelength is sought. The relationship between the wavelength λ and the cursor position C is thus given by

$$\lambda = A_0 + A_1 (C-1) \qquad (1)$$

where A_0 and A_1 are constants.

In practice the arc jet is turned and set at the appropriate power level and the mass flow rate. The radiant energy from the arc jet is incident upon the slit of the spectrograph, is dispersed by the spectrographic optics and is then analyzed by the OMA. An X-Y recorder attached to the OMA makes it possible to obtain a hard copy of the spectrum.

The first task in this project was to determine the composition of the free flow at different operating conditions. In Table I, test conditions used are given.

TABLE I

Arc Jet Test Conditions Used to Obtain Spectra of the Free Stream

Flow Rate (lb-mass/sec*)	Current (A)	Power (MW)	Enthalpy Btu/lb**
0.05	500	0.87	7,000
0.05	1000	1.76	10,500
0.05	1800	2.76	13,000
0.10	1000	2.38	9,800
0.30	500	2.34	5,000

* To convert to SI unit of Kg/Sec multiply by 0.454

** To convert to SI unit of MJ/Kg multiply by 2.32×10^{-3}

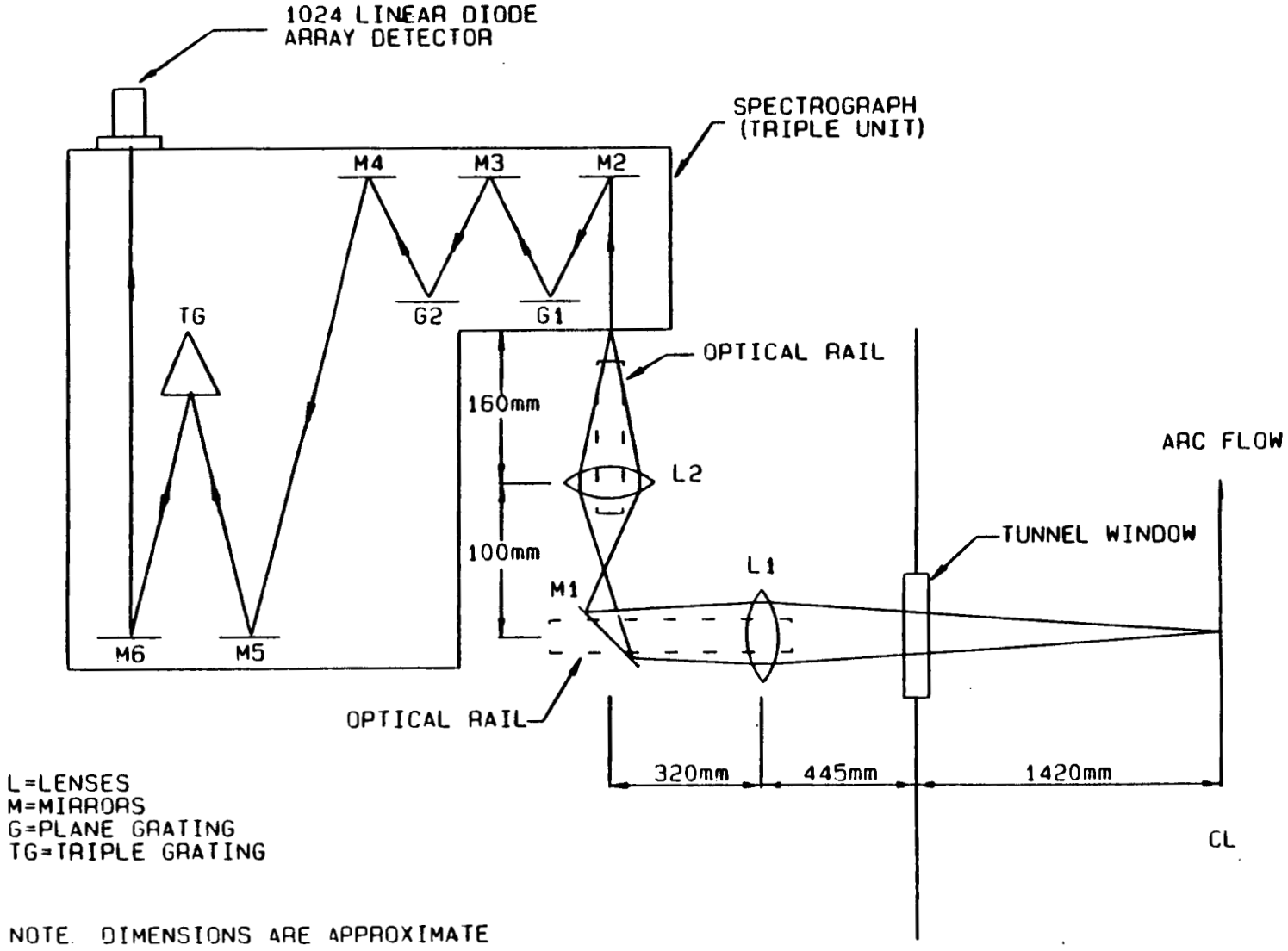


Figure 1.- Spectrograph side optical train (not to scale).

RESULTS AND DISCUSSION OF SPECTRA OBTAINED

It appeared at the very outset that a spectrum obtained depended more on the enthalpy than on the power setting. At low enthalpy settings i.e., at both 5000 Btu/lb and at 7000 Btu/lb, the low wavelength part of the spectrum was totally unreadable, because the signal to noise ratio was very poor. In general the quality of the spectrum was bad and the band heads were not clearly defined. At the enthalpy of 5000 Btu/lb (about 3.85 eV per O_2 or N_2 molecule), the probability of dissociation of either molecule is small. This is borne out by the fact that the majority species discovered in the spectrum at that setting was neutral N_2 molecule and some O_2 molecular bands. One atomic oxygen line was observed at 700.2 nm. The identification of all species were done with the aid of a) the MIT Wavelength Tables (6), b) Pearses' "Handbook on Identification of Molecular Spectra" (7), c) Herzberg's "Atomic Spectra" (8). Plates recorded by Willey (9) were also consulted in the process to aid in the identification process.

Since the catalog of spectral information recorded and species identified is voluminous, it is being compiled into an appendum to this report and is available by contacting John Grimaud at the JSC arc jet facility.

The high enthalpy conditions gave some extremely interesting results. In the highest enthalpy condition that was run it was discovered that entire low wavelength side of the spectrum was inundated with tungsten spectral lines. The cathode of the arc is made of a tungsten alloy, but that fact on its own does not explain the abundance of highly intense tungsten lines throughout the spectrum. Tungsten burns in air with a white flame, however, an observer standing near the arc jet during these measurements reported seeing a yellow radiation from the arc jet. This discounts to a large extent the notion that a large piece of the tungsten electrode may have become dislodged from the cathode and was burning in the stream. Besides the tungsten, nitrogen, oxygen, copper, and thorium lines were found. Examples of spectra recorded are shown in the accompanying figures (Figs. 2-4).

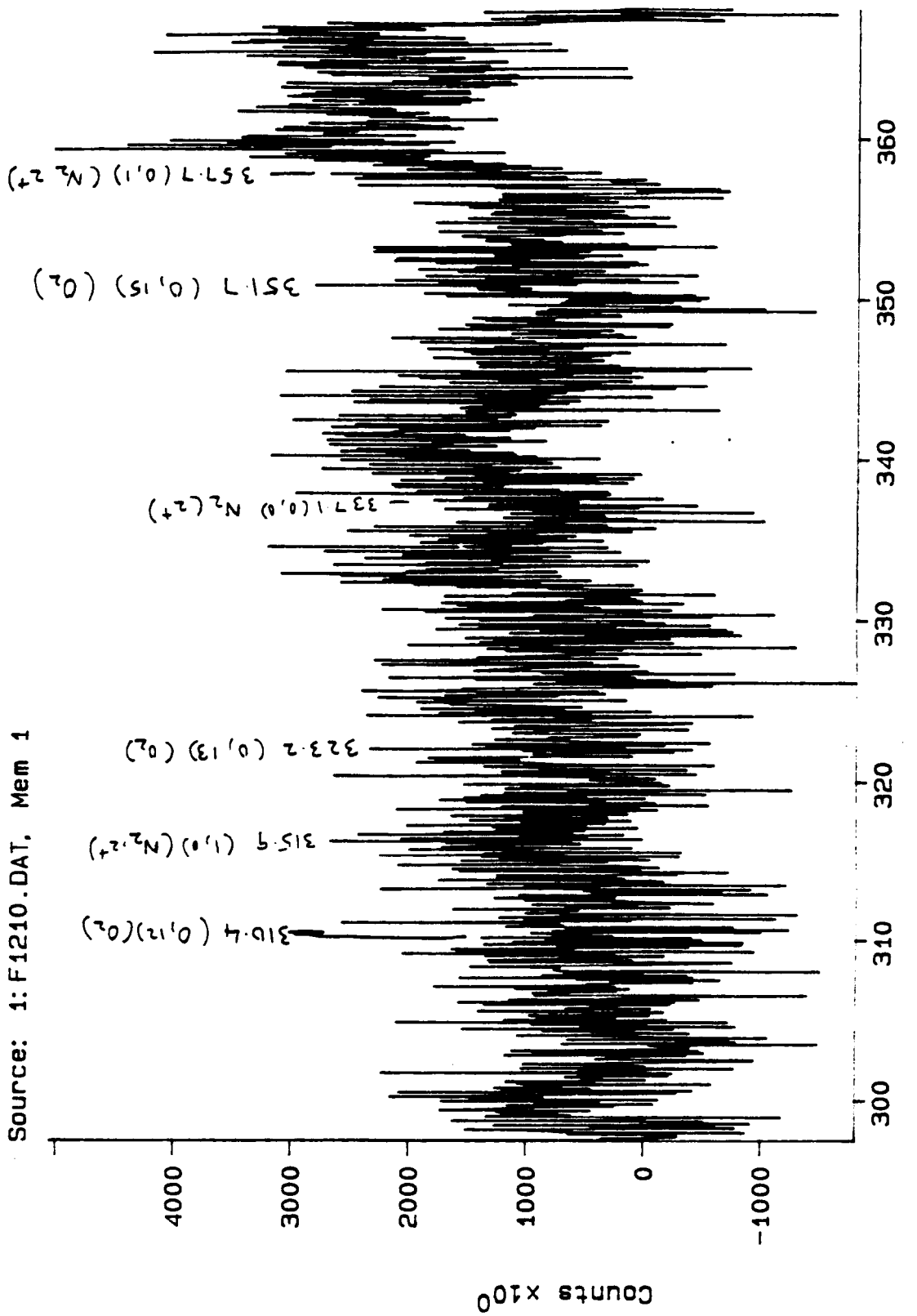


Figure 2.- Part of the spectrum of the free stream recorded at a low enthalpy condition, showing N_2 and O_2 bands. Notice the poor quality of the spectrum and the band heads.

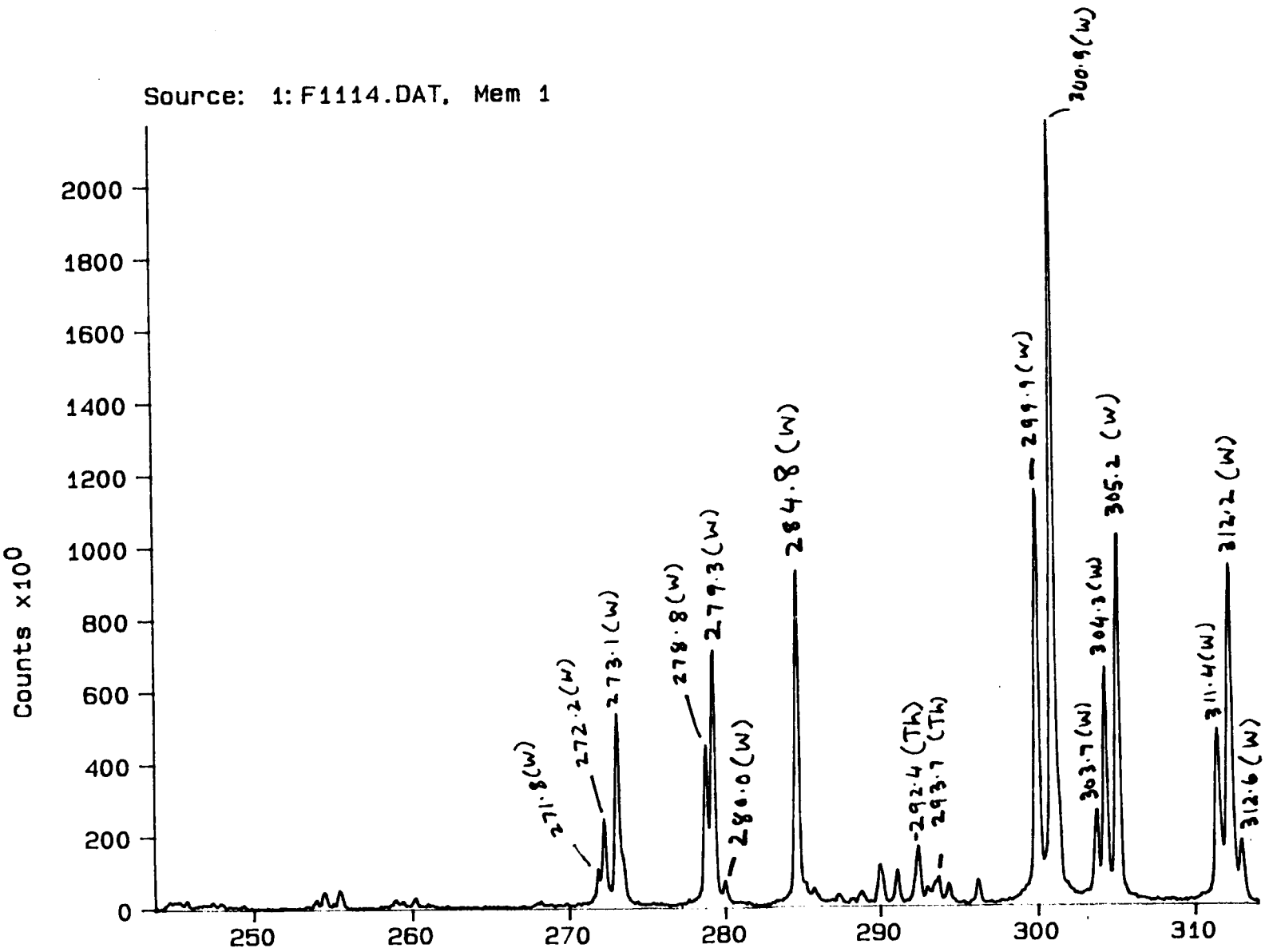


Figure 3.- Part of the free stream spectrum taken at an enthalpy of about 1000 BTU/lb, showing the lines due to tungsten.

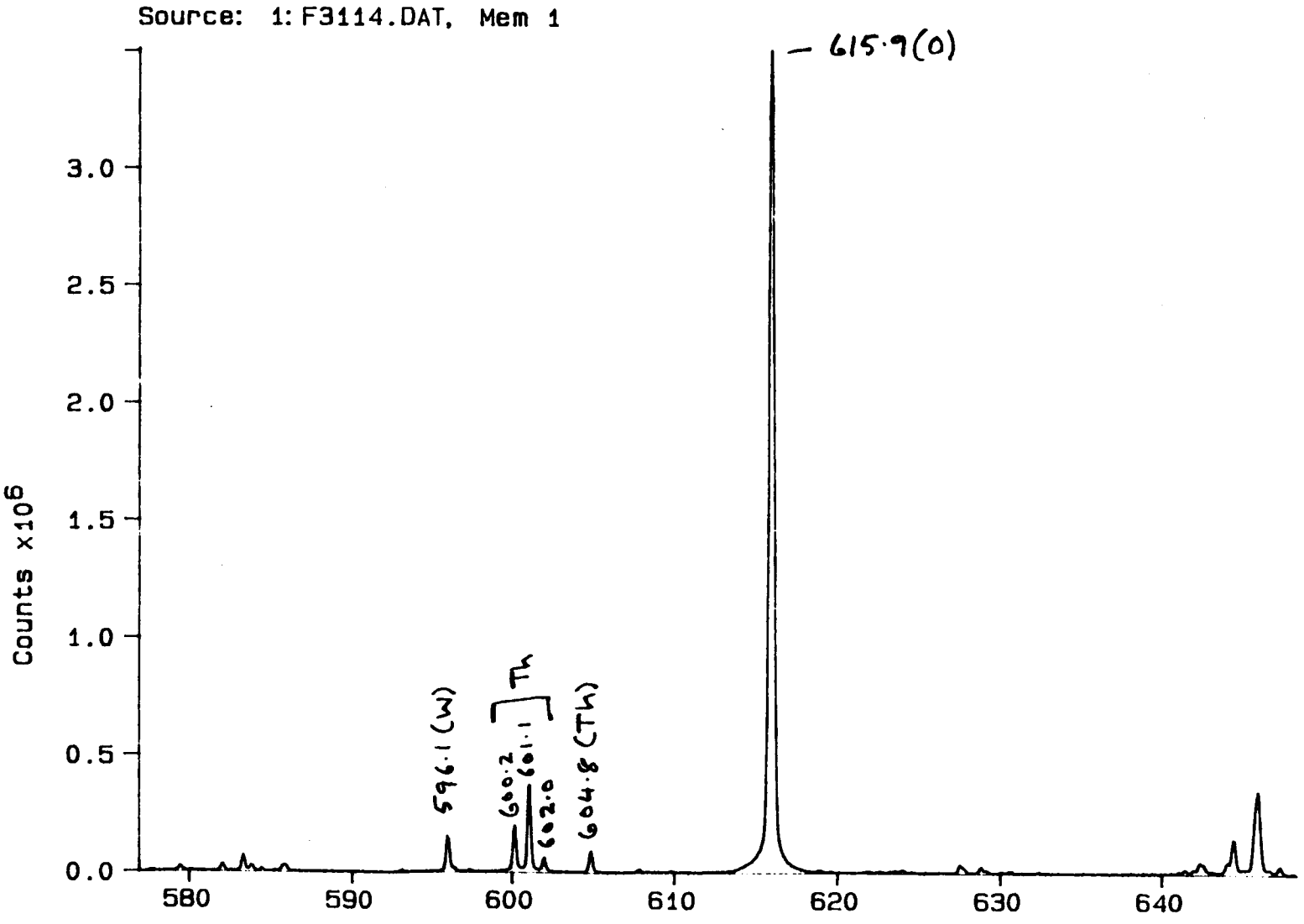


Figure 4.- Another part of the free stream spectrum taken at high enthalpy conditions, showing thorium lines and an oxygen line.

LASER DIAGNOSTICS

Since the advent of the high powered laser (both continuous wave and pulsed), they have come into use increasingly as diagnostic probes (10). The list of the laser techniques which have come into vogue is a long one. There are four major techniques which should be enumerated here.

1. Spontaneous Raman Spectroscopy (SRS). A technique which allows species identification, concentration measurement and measurement of temperatures.

2. Laser Induced Fluorescence (LIF). A technique by which species identification, concentrations and temperatures can be measured. The technique, however, can be used only for a handful of molecules and free radicals, most of which are reaction intermediates.

3. Coherent Anti-Stokes Raman Spectroscopy (CARS). A superior technique for the measurement of temperature.

4. Laser Doppler Velocimetry (LDV). A technique used for determining particle velocities and from which the size of the particles can be inferred indirectly.

The laser methods in general have the following advantages.

- Non-intrusive (no probes, thus capable of penetrating and investigating hostile environments such as the arc jet)
- High spatial resolution ($\sim 10^{-3}$ mm^3)
- High temporal resolution (10^{-8} sec)
- High spectral resolution ($\sim 10^{-2}$ nm)
- High sensitivity ($\sim$ ppm)
- Multiplex advantages (simultaneous measurements of several species at one point or of one species at several points)

It therefore behooves us to attempt one or more of the above mentioned techniques in the study of the free flow and/or shock layer in the arc jet. It was decided by the arc-jet facility coordinators that SRS should be the first technique that should be tried. A schematic of the SRS phenomenon is shown in Fig. 5. As can be seen from this figure an atom, molecule or free radical is excited to a virtual state and the transition back to the lower states results in scattering with one component on the higher wavelength of the incident radiation and one component on the lower wavelength side of the excitation radiation (11).

Classical Theory of Raman Effect. In order to simplify this classical derivation (11) of the Raman Effect, consider a diatomic molecule of polarizability α_Z. Thus an electric field E_Z along the Z axis would result in a dipole moment $\mu_Z = \alpha_Z E_Z$. The molecular polarizability (unlike in the case of an atom) depends on the internuclear separation and also on the angle between the molecular axis and the electric field. If the normal frequency of vibration is ν_{Vib} and r is the vibration amplitude then to a first order approximation,

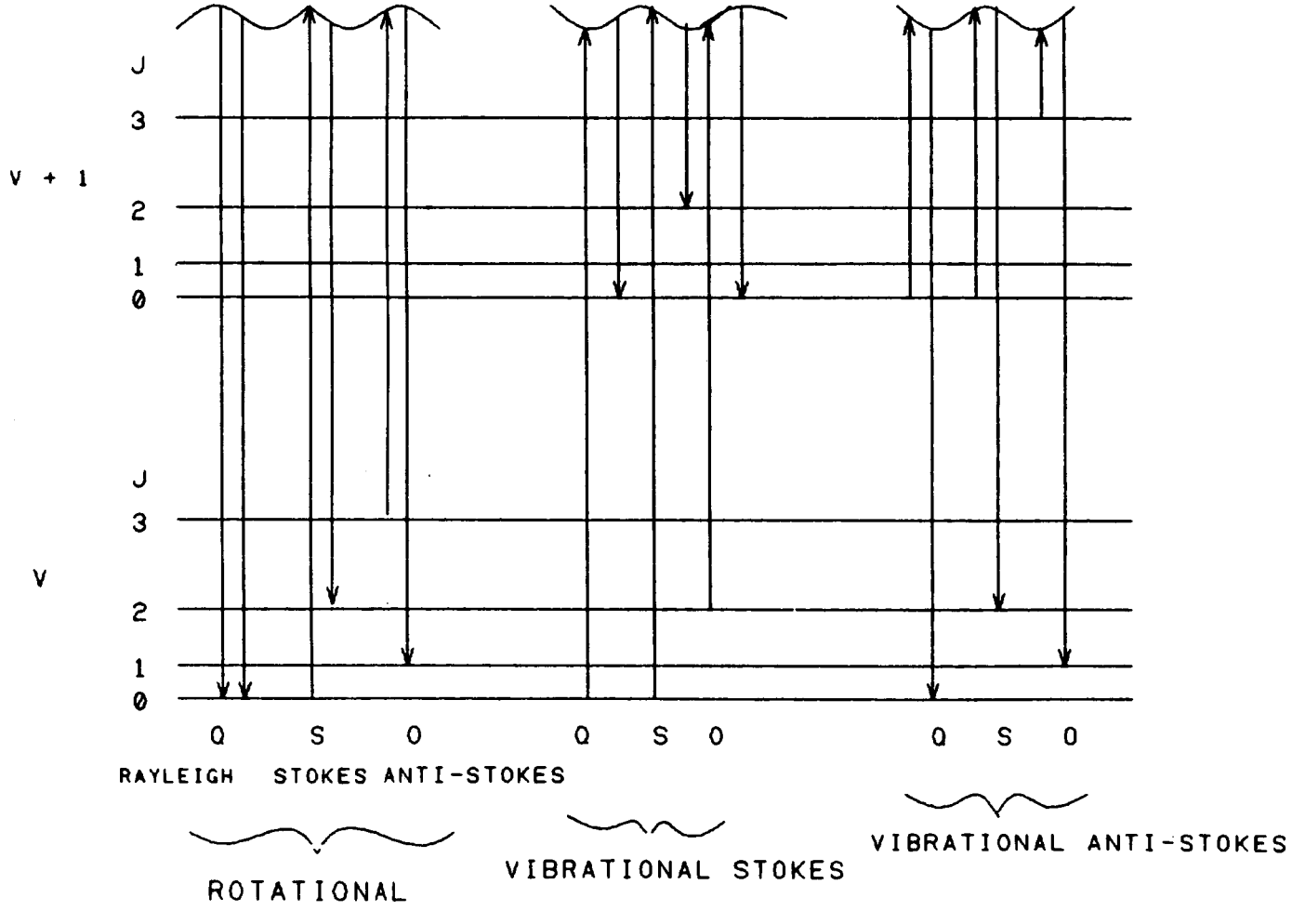


Figure 5.- Rotational and vibrational-rotational Raman scattering processes.

$$\alpha_z = \alpha_{oz} + \gamma \frac{\partial \alpha}{\partial R} \cos 2\pi \nu_{Vib} - t \qquad (2)$$

In other words the polarizability fluctuates sinusoidally about a mean value of α_{OZ}. In an electric field E_Z of radiation of frequency ν_0

$$E_z = E_o \cos 2\pi \nu_o t \qquad (3)$$

Combining the two equations we get

$$\mu_z = E_z \alpha_z = E_{oz} \cos 2\pi \nu_o t + E_{oz} \gamma \frac{d\alpha}{dR} \cos 2\pi \nu_{Vib} t \cos 2\pi \nu_o t \qquad (4)$$

The first term in the above equation in this classical model gives rise to the elastic Rayleigh scattering. Using a well known trigonometric identity, the second term can be rewritten as

$$\frac{1}{2} E_{oz} \gamma \frac{\partial \alpha}{\partial R} \cos 2\pi (\nu_o + \nu_{Vib}) t + \frac{1}{2} E_o \gamma \frac{d\alpha}{dR} \cos 2\pi (\nu_o - \nu_{Vib}) t \qquad (4')$$

Thus the classical theory predicts that in addition to Rayleigh Scattering there will be radiation with frequencies $\nu_o \pm \nu_{Vib}$ and we identify them with the observed anti-Stokes and Stokes lines of Raman Scattering.

Laser Raman Spectroscopy In the Arc Jet.

A 5 Watt laser (Spectra Physics Series 2000 Argon/ion laser) was chosen for setting up the Raman Scattering Experiment. A series of high reflectance mirrors and high quality focusing lenses were used in order to guide the laser beam into the arc jet. The optical arrangement and the laser beam path are shown in Fig. 6 and Fig. 7. The beam scattered at right angles to the incident beam exits the tunnel and is then analyzed by the spectrographic set up. At the time of writing of this report, no experimental data was obtained.

FURTHER WORK TO BE ACCOMPLISHED.

1. Recently the vibrational temperature (12) in the shock layer was obtained by examining the vibrational bands of the first negative system of singly ionized nitrogen molecule. Remarkable agreement was obtained between the experimental results and results obtained from the computer code NEQAIR (13). This computer code plots a theoretical spectrum when certain parameters (temperature included) are input.

The next obvious step is to determine the rotational temperature in the shock layer by examining the rotational lines of certain bands of N_2^+ (first negative system). The technique used has been described in textbooks (14). An outline of the procedure to be followed to obtain the rotational temperature has been chalked out and it is expected that it will be followed up.

2. At the time of writing of this report, the necessary optics was partially in place for the Raman set-up. The laser was turned on and it operated satisfactorily. It is expected that the optics will be completely in place

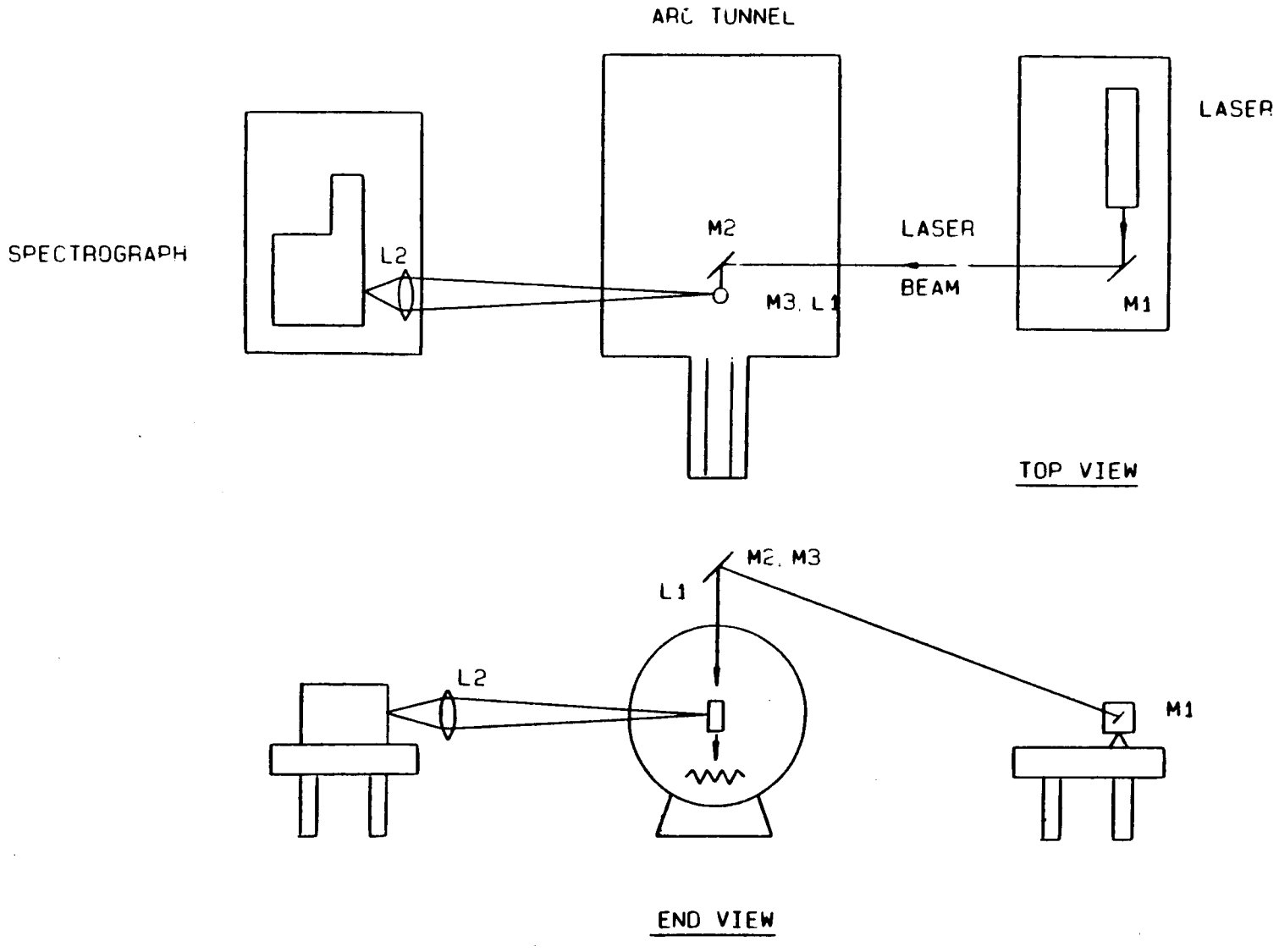


Figure 6.- Spectrograph optical path schematic (not to scale).

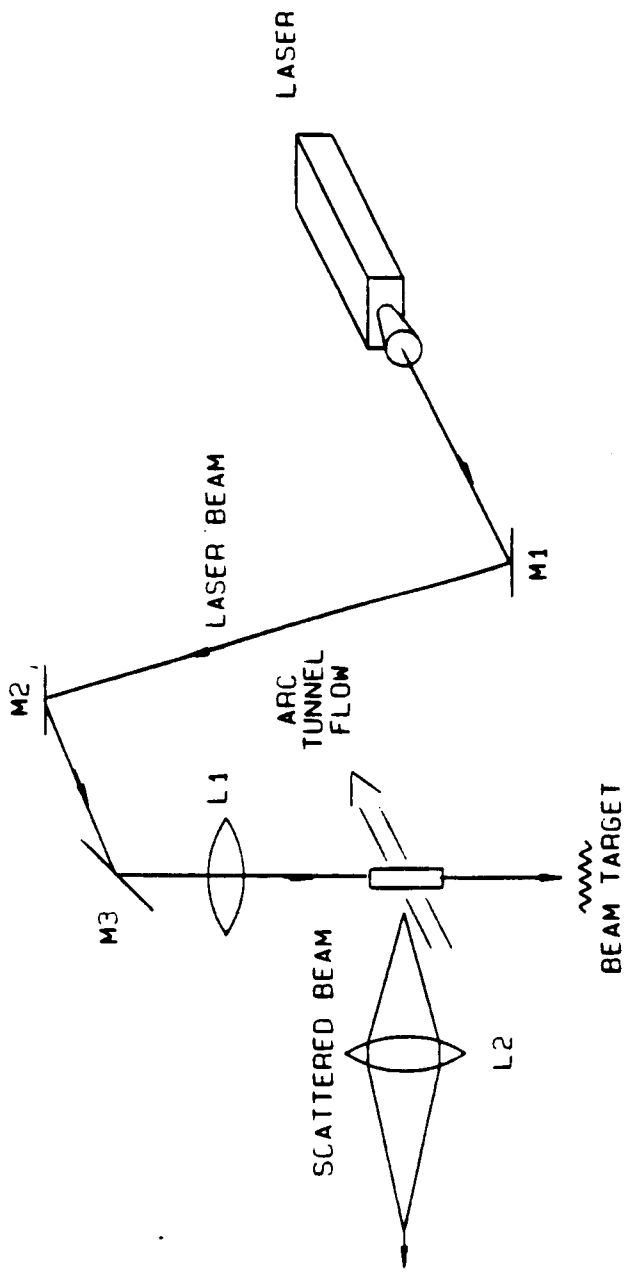


Figure 7.- Three-dimensional path.

anytime now. One of the first experiments scheduled is to determine the Raman shifts of NO, N_2 and O_2 in the free stream. Table II shows the results of a simple, calculation giving the location of the Stokes and anti-Stokes line for some of the excitation wavelengths obtainable from the laser.

TABLE II

Argon Ion laser λ = 514.5 nm

Species	Raman Shift cm^{-1}	$\Delta\lambda$ (nm)	Stokes (nm)	Anti-Stokes (nm)
O	226	6.0	520.5	508.5
N_2	2331	61.7	576.2	452.8
NO	1876	49.7	564.2	464.8
O_2	1556	41.2	555.7	473.3
		λ = 476.5 nm		
O	226	5.1	481.6	471.4
N_2	2331	52.9	529.4	423.6
NO	1876	42.6	519.1	433.9
O_2	1556	35.3	511.8	441.2
		λ = 454.5 nm		
O	226	4.7	459.2	449.8
N_2	2331	48.2	502.7	406.3
NO	1876	38.8	493.3	415.7
O_2	1556	32.1	486.6	422.1

In real applications it is necessary to consider the effect of the rotational energy levels upon the observed scattered spectrum. For diatomic molecules, the selection rules for radiative scattering are $\Delta V = 0, \pm 1$ and $\Delta J = 0, \pm 2$, where V is the vibrational quantum number, and J is the rotational quantum number. Of the three branches $\Delta J = 0$ (Q branch) is brighter than the other two namely $\Delta J = -2$(O) branch and $\Delta J = 2$ the S branch.

If the Stokes and anti-Stokes intensities can be measured in a medium, which is in local thermodynamic equilibrium, then the temperature can be uniquely determined from the ratio of these measurements. The ratio of a Stokes and an anti-Stokes lines corresponding to the same J value in the Q Branch is given by (15)

$$\frac{I_{STOKES}}{I_{ANTI\text{-}STOKES}} = \left[\frac{\nu_0 - \frac{\Delta E}{h}}{\nu_0 + \frac{\Delta E}{h}}\right]^4 e^{\Delta E/KT}$$

where ν_o is the frequency of the incident radiation and Δ E is the energy of the rotational lines assumed constant for all lines in the Q branch. This might be difficult in the arc jet where thermal equilibrium is difficult to achieve. Perhaps in the shock layer this condition might be met and the method might be successful in shock layer temperature determination.

Other techniques have been used to determine temperature and species concentration in hypersonic flows using rotational Raman spectra (16). These techniques need to be examined to see if they can be applied to the arc jet conditions.

3. There have been suggestions made by professionals here that the flow velocity be determined experimentally. A knowledge of the flow velocity would yield other parameters of the flow. At present an in-depth study of this problem needs to be undertaken and a clear-cut procedure established and demonstrated.

REFERENCES

1. Scott, C. D., "Catalytic Recombination and Space Shuttle Heating," AIAA Paper 82-0481, 1982.

2. Scott, C. D., Ried, R. C., Maraia, R. J., Li, C. P., and Derry, S. M., "An AOTV Aeroheating and Thermal Protection Study," AIAA Paper 84-1710, 1984.

3. Scott, C. D., "Effects of Nonequilibrium and Wall Catalysis on Shuttle Heat Transfer," J. Spacecraft and Rockets, Vol. 22, No. 5, pp. 489-499.

4. Scott, C. D., "Line Broadening of the Triplet Diffuse Series of Helium in an Arc Jet Plasma," Phd Dissertation, University of Texas at Austin, 1969.

5. Blackwell, H. E., Scott, C. D., Hoffman, J. A., Mendez, S. B., and Swenson, G. R., "Spectral Measurements of the Space Shuttle Leeside Shock Layer and Wake," AIAA Paper 86-1262, 1986.

6. MIT Wavelength Tables, The M.I.T. Press, Cambridge, Mass., 1963.

7. Pearse, R. W. B. and Gaydon, A. G., "The Identification of Molecular Spectra, Chapman and Hall, London, 1965.

8. Herzberg, G., Atomic Spectra and Atomic Structure," Dover, New York, 1944.

9. Willey, Ronald J., "The Identification of Excited Species in Arc Jet Flow," Final Report, Summer Faculty Fellowship Program, NASA-JSC, 1984.

10. Murty, V. S., "Laser Diagnostics in Combustion" Final Report, ORAU Summer Faculty Program, U.S. Department of Energy, Pittsburgh Energy Technology Center, Pittsburgh, PA, 1986.

11. Herzberg, G., "Molecular Spectra and Molecular Structure, Vol. II Infrared and Raman Spectra of Polyatomic Molecules," D. Van Nostrand Co., Inc., New York, N.Y. (1945).

12. Blackwell, Harvel E., Wierum, F. A., and Scott, C. D., "Spectral Determinations of Nitrogen Vibrational Temperatures," AIAA Paper 87-1532, 1987.

13. Park, Chul, "Nonequilibrium Air radiation (NEQAIR) Program: Users Manual", NASA TM-86707, 1985.

14. Herzberg, G., "Molecular Spectra and Molecular Structure, Vol. I Spectra of Diatomic Molecules", D. Van Nostrand Co., Inc., New York, N.Y. (1950).

15. Schreiber, P. W., Gupta, R., and Roh, W. B., "Application of Lasers to Combustion Diagnostics," SPIE Vol. 158 Laser Spectroscopy (1978).

16. Lewis, J. W. L. and Williams, W. D., "Measurement of Temperature and Number Density in Hypersonic Flow Fields Using Laser Raman Spectroscopy" AIAA Paper 75-175, 1975.

ACKNOWLEDGMENTS

I wish to thank NASA and ASEE for the financial support given to me for this work. I wish to thank my JSC colleagues, John Grimaud and Carl Scott, for their guidance and support throughout this work. I extend my sentiments of appreciation to my fellow Summer Faculty Fellow, Ronald Willey, and also to Fred Wierum, who taught me the various aspects of the facility. Finally, my thanks to all of NASA and the Lockheed personnel of the arc jet facility who made my stay here very pleasant and rewarding.

PETROGENESIS OF BASALTS FROM THE ARCHEAN MATACHEWAN DIKE SWARM SUPERIOR PROVINCE OF CANADA

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Dennis O. Nelson, Ph.D.
Academic Rank:	Professor
University & Department:	Sul Ross State University Department of Geology Alpine, Texas 79832

NASA/JSC

Directorate:	Space and Life Sciences
Division:	Solar System Exploration
Branch:	Experimental Planetology
JSC Colleague:	Don A. Morrison, Ph.D.
Date:	August 21, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

The Matachewan Dike swarm of eastern Ontario comprises Archean age basalts that have been emplaced in the greenstone, granite-greenstone and metasedimentary terrains of the Superior Province of Canada. The basalts are Fe-rich tholeiites, characterized by the near ubiquitos presence of large, compositionally uniform, calcic plagioclase (AN 85±5). Major and trace element whole-rock compositions, along with microprobe analyses of constituent phases, from a group of dikes from the eastern portion of the province, were evaluated to constrain petrological processes that operated during the formation and evolution of the magmas. Three compositional groupings, have been identified within the dikes. One group has compositional characteristics similar to modern abyssal tholeiites, e.g. light-REE depleted, and is termed morb-type here. A second group, enriched in incompatible elements and light-REE enriched, $(La/Sm)n > 1.80$, is referred to as the enriched group. The third more populated group has intermediate characteristics and is termed the main group. Owing to this latter group's larger and more coherent compositional variation, geochemical modeling was concentrated here. Although some dikes of the main group are relatively uniform, others display significant variation on the outcrop scale. Geochemical trends of the intradike suite are parallel to those of the interdike suite and are considered together. Major element mixing models suggest that the compositional variation of the main group can be accommodated by fractionation of assemblages consisting of variable proportions of olivine, clinopyroxene and plagioclase. Trace elements, however, cannot be accommodated at the same fractionation level as the major elements, indicating that processes more complex than simple crystal fractionation must have operated during main group evolution. The open-system process of periodic replenishment can explain both the uniform composition of the plagioclase megacrysts as well as the apparent decoupling of the major and trace element compositions. Combined replenishment-fractionation, however, cannot account for the compositional variations between the three groups of the dikes. The observation of both morb-type and enriched compositions within a single dike of the SEC strongly argues for the contemporaneous existence of magmas derived through different processes. Mixing calculations suggest that two possibilities exist. The least evolved basalts of the main group lie on a mixing line between the morb-type and enriched group, suggesting mixing of magmas derived from heterogeneous mantle. Alternatively, mixing of magmas derived from a depleted mantle with heterogeneous Archean crust can duplicate certain aspects of the Matachewan dike compositional array.

INTRODUCTION

Basaltic rocks constitute a significant portion of the volcanic rocks of the Archean crust. They occur as flow rocks within greenstone and greenstone-granite terrains, and are compositionally similar to temporally and spatially related mafic intrusions. In addition, basalts comprise large Archean dike swarms, the areal extent of which results in the dikes crossing litho-tectonic boundaries along which Archean blocks of markedly different character are juxtaposed. Interest in Archean basalts stems from the information they might provide relative to crustal genesis and plate tectonic processes in the early Precambrian.

This particular study concerns the Matachewan dike swarm (MS) of eastern Ontario. This swarm was emplaced over an area exceeding 250,000 km2 in a stable cratonic "granitic"-greenstone terrain of probable continental origin (1,2). The dikes are tholeiitic and cross lithologies varying from mafic to felsic volcanics, volcaniclastics and intermediate to silicic plutonic bodies (3,4). The majority of the MS dikes host megacrysts of plagioclase of uniform composition (AN 85±5), similar to constituent plagioclase in Archean anorthosites (2,5,6). As a result of this megacryst connection, a geochemical study of a portion of the MS dikes was undertaken to determine the magmatic processes involved in basalt production and evolution with the ultimate goal of constraining anorthosite genesis. In this paper, we report the results of geochemical modeling of the basalts and the resulting implications for their open-system evolution.

GEOLOGIC SETTING

The Matachewan dike swarm occurs within the Superior Province of eastern Ontario (fig. 1), crossing several subprovinces (7). These subprovinces are characterized as greenstone-tonalite ("granite"), paragneiss and batholithic terrains, differing in the proportions of different rock types, metamorphic grade and structural style (8). The subprovinces can be described (7, 9) as metavolcanic-rich belts (e.g. the Abitibi, Wawa and Wabigoon subprovinces) separated by intervening metasedimentary belts (e.g. the Quetico subprovince), although the boundaries between some belts are tectonic in nature (5). Plutonic rocks are as diverse as the volcanic (3, 7), although tonalite-granodiorite and monzonite-quartz monzonite appear to dominate (4). Plutons as evolved as granite, as well as syenites, are found (7). Layered mafic intrusions, often with associated anorthosites, also occur within the Superior Province (10, 11).

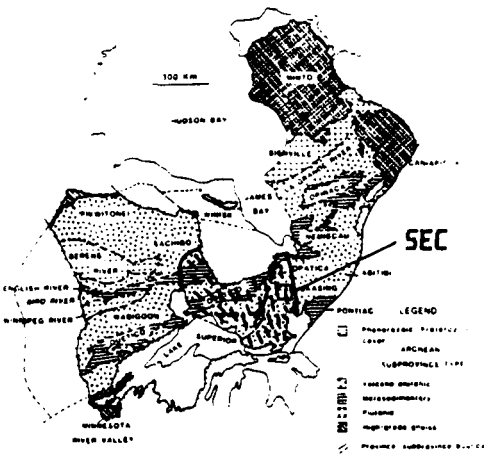


Figure 1. Generalized geologic map of the Superior Province (from Card(7)). Trend and distribution of the Matachewan dike swarm is outlined and the southeast cluster (SEC), emphasized in this study, is indicated.

The ages of magmatic, metamorphic and tectonic events were broadly synchronous in the Superior Province but initiated somewhat earlier in the northern subprovinces (7). Within the Wabigoon, Wawa and Abitibi belts (Fig. 1), U-Pb zircon dates indicate that activity occurred over the period of 2790-2660 Ma (7, 9, 10), and can often discern an early volcanic-plutonic event, a middle tectonic-metamorphic event and a late plutonic event (7, 9). The Matachewan dikes were emplaced after the major periods of magmato-tectonic activity (12) in the approximate time frame of 2.5 to 2.6 Ga (13). The subparallel trend of these dikes suggests emplacement during extensional tectonics.

SAMPLE DISTRIBUTION AND CHARACTER

Geochemical modeling was concentrated on samples from a portion of the MS, the southeast cluster (SEC) of figure 1, because of higher sample density and greater compositional variability. Most of the SEC dikes contain large (up to 20 cm) plagioclase phenocrysts (2). These phenocrysts are often irregularly distributed (1) and flow differentiation processes are evident in outcrop. Recognizing non-liquid compositions, produced by cumulate processes, on geochemical variation diagrams is often difficult. Frequently, more than one subparallel liquid descent line is present producing an apparent scatter. Europium anomalies (Eu/Eu^*), which monitor the participation of feldspar in magmatic evolution, are unequivocal as an indication of

plagioclase accumulation in basalts only when Eu/Eu* > 1.0. If the basalts have evolved a negative Eu anomaly (i.e. Eu/Eu* < 1.0), subsequent plagioclase accumulation may occur without a resulting positive Eu anomaly, while producing a trend similar to crystal fractionation. Major elements, however, are controlled by phase equilibria. Significant departures from evolution lines may therefore indicate compositions that were never liquids. On a CaO-MgO diagram, the bulk of the MS dikes define a single linear array. Several, however, plot off this trend, often along a line towards either olivine or plagioclase. Because it is suspected that these samples reflect the presence of cumulus minerals, these have not been considered in the modeling procedures below.

Modification of original magmatic concentrations through metamorphism, of concern here for the MS dikes, has been addressed by Pearce and Cann (14) and many others. In the SEC data, strong correlations exist between the immobile elements (e.g. Ti, Zr and Y) and most other elements. Exceptions are Rb, K, Ba and Sr, which, although showing considerable internal consistency, are scattered when plotted against Zr. These four elements are regarded to have been mobile during metamorphism and are not used in modeling, although it is suspected that the ratios of these components reflect magmatic values.

GEOCHEMISTRY

The MS basalts vary from olivine- to quartz-normative tholeiites. When plotted on the tectono-magmatic diagrams of Pearce and Cann (14), the MS rocks fall in the field of ocean-floor basalts, a characteristic of many basalts of greenstone belts (15). For the MS dikes, ratios K/Rb, K/Ba and Sr/Rb are more comparable to Type II than Type I basalts. In terms of the rare earth elements (REEs), the MS dikes vary from light-REE depleted to light-REE enriched (Fig. 2). In the latter case, the MS dikes are more similar to continental flood basalts (16) or oceanic island tholeiites and alkali basalts (16, 17) than abyssal tholeiites. The MS compositions reflect an Fe-enrichment trend (2); The MgO content of the least evolved SEC compositions is near the density minimum for tholeiitic magmas (18, 19).

The MS data define linear arrays in compositional space (e.g. Fig. 3) indicating that the magma(s) evolved through similar processes. Examination of figures 2 and 3 allow the delineation of three compositional groups within the SEC. The first, the morb-type, has low incompatible element concentrations (Fig. 3) and is light-REE depleted (Fig. 2), similar to abyssal tholeiites. A second, or enriched

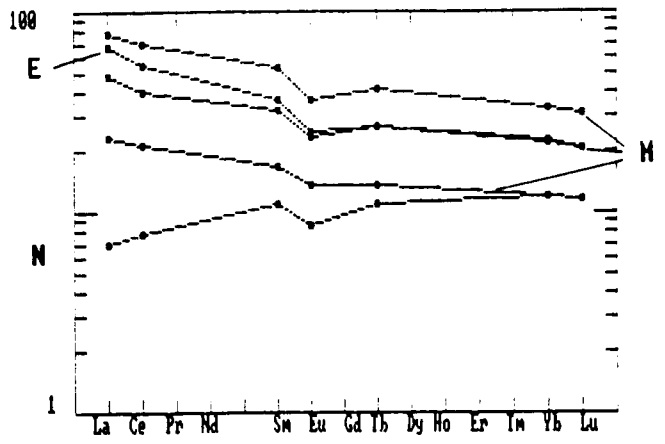


Figure 2. Rare earth element diagram for selected rocks of the SEC group. The symbol N denotes Morb-type; E the enriched group; M the main group.

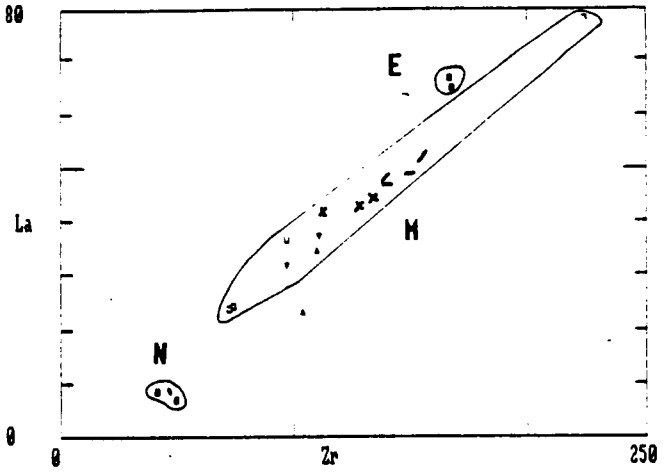


Figure 3. Zr versus La for the SEC rocks. Group designations as in figure 2. Common data symbols on the graph represent samples from the same dike.

group, is light-REE enriched (Fig. 2) and has high incompatible element concentrations (Fig. 3). The third, or main group, comprising the bulk of the SEC, has intermediate compositional characteristics.

Within the SEC, both intra- and interdike variations are found (Fig. 3). Intradike suites, collected at the same outcrop for a given dike, range from uniform to rather varied. The trajectories of the intradike variations are, however, subparallel to those of the interdike array, and are considered to have evolved through similar processes.

Two samples in figure 3 (+ symbols) are not dikes, rather they are older basaltic flow rocks. These flows are very similar to the morb-type dikes with the exception that the flows are less evolved (e.g. higher Mg numbers, Cr and Ni contents). It is suspected that magmas similar to the flow rocks were produced during Matachewan time. Accordingly, flow rock compositions have been used as parental liquids in some of the models below.

DISCUSSION

Linear compositional trends may result from variations imposed by the source, by the melting event, by chamber processes, by mixing, through crustal interaction and as a result of post-emplacement processes, acting independently or together. It is the purpose of the modeling here to determine: 1) what processes operated, 2) their relative contributions to the observed variations, and 3) when in evolution scheme the individual processes acted.

RESIDENCE CONTAMINATION

"Residence contamination" is the post-consolidation exchange of components between the wall-rock and the dike (20). Sample density in the present study is not sufficient to detect the compositional gradients in the dikes, nor is wall-rock data available in all cases. As a result, the role of residence contamination in the SEC could be quantitatively evaluated on only a few dikes. The compositions of the interiors and margins of relatively uniform dikes emplaced in different terrains were examined for compositional shifts that could be attributed to wall-rock exchange. No such shifts were observed. For example, dikes intrusive into "granitic" material were not enriched at the margins in elements such as Zr, La, Th and Y. Dikes intrusive into mafic terrains do not reflect enrichments of Ni, Sc, Co and Cr. One dike, emplaced along a granite-amphibolite contact was particularly useful. No compositional difference was detected between the interior and either margin (Fig. 4). On a regional scale, with the exception that the morb-type dikes appear to occur only in greenstone belts, no significant differences are apparent between those dikes emplaced in the greenstone versus "granitic" terrains (Fig. 5). Although this does not preclude residence contamination, it clearly limits its role to a minor one in the evolution of the main SEC group.

FRACTIONAL CRYSTALLIZATION

Fractional crystallization (FC) has undoubtedly operated

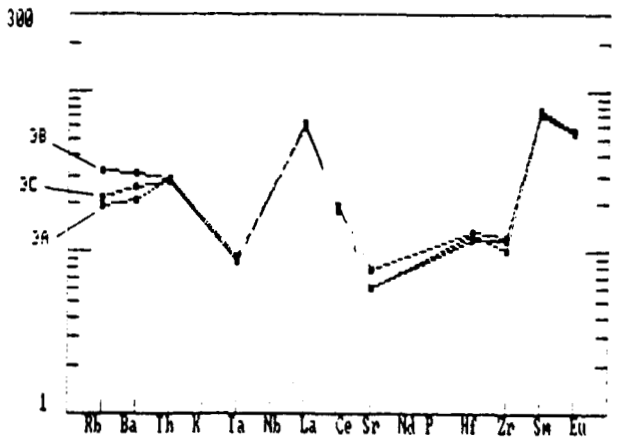


Figure 4. Spider diagram for a single intradike suite. Sample 3B is from the dike interior, 3A and 3C are adjacent to granite and amphibolite, respectively.

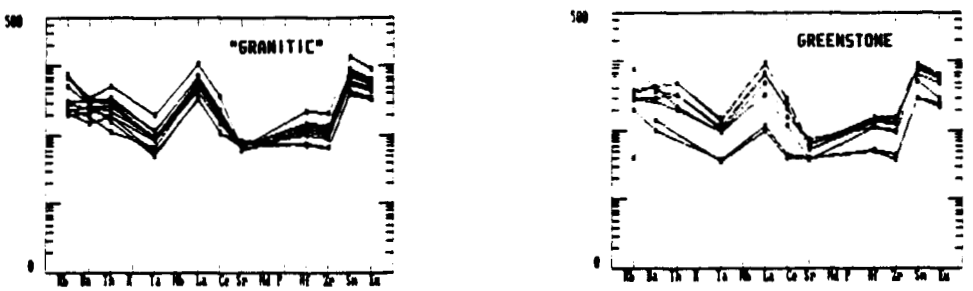


Figure 5. Spider diagram comparing compositions of SEC dikes emplaced in granitic and greenstone terrains.

at some level to produce the evolved SEC compositions. This is indicated by low Mg numbers (40.7-57.1), Cr (42-101) and Ni (40-80) contents and the Fe-enrichment trend. Additionally, the MS dikes plot along the 3-5 Kb plagioclase-olivine-clinopyroxene cotectic (21) as well as defining an array in Al2O3-CaO space along a plagioclase-clinopyroxene control line. It is important to note, however that the large variation in (La/Sm)n between the morb-type (0.62-0.74), the main- (1.39-1.58) and the enriched- (1.81-2.41) groups preclude FC from being the control of the intergroup

compositional differences of the SEC. Simple FC cannot produce significant changes in this ratio. Covariations between Zr and Y, P, and La, and between La and Yb limit the roles of phases that can fractionate La and Sm, e.g. garnet, apatite, zircon and monazite.

To evaluate FC in the SEC main group, major element mixing models were performed using phenocryst and whole rock data of supposed evolved and parental compositions. In general, the lowest residuals were achieved for assemblages of olivine, clinopyroxene and plagioclase. Magnetite was not considered because it is not a phenocryst phase, because Ti is incompatible and because of the marked Fe-enrichment and lack of SiO2 enrichment in the series (22). Successful models were achieved in some, but not all, members of both the intra- and interdike suites. In most dikes, the marginal samples were less evolved than those of the interior. In some dikes, however, the reverse is true.

The calculated assemblages are capable of producing the trace element variations observed, but at markedly different F values than that required by the major elements. In particular, the incompatible elements require that the F value be smaller, e.g. the fractionation greater, than that predicted by the mixing models. This problem cannot be overcome by assuming different partition coefficients (Kd). Calculated bulk Kds in the assemblages are already low (e.g. < 0.05-0.15) for these elements. To produce the observed variation within the restrictions of F required by the mixing models, the Kd values must be negative, varying for example, from -0.32 for Zr to -0.67 for La, an impossible situation. The decoupling of the major and trace elements indicates that simple FC is not capable of accounting for the compositional variability of the main SEC group.

The failure of the FC model suggests that the compositional variation observed within single dikes was not the result of insitu processes. This suggests that the composition of the magma emplaced in the dike changed with time. The dikes should not, therefore, be viewed as the result of "instantaneous" emplacement of a homogeneous magma. Rather, they record a history of evolution that occurred prior to dike formation.

IMPLICATIONS OF THE PLAGIOCLASE MEGACRYSTS

Constraints to petrogenetic models for the SEC main group are provided by the plagioclase megacrysts. A compositional variation of less than three AN units within a single cryst requires growth in an essentially isothermal, compositionally uniform environment (2, 23). This composi-

tional uniformity extends to all megacrysts within the MS dikes, where AN contents are 85±5 regardless of the dike's geologic setting and trace element concentration levels (2). Further, the observed partitioning of REE between plagioclase megacrysts and matrix indicates equilibrium between megacryst and host dike (2, 23).

These observations indicate that the megacrysts grew in equilibrium in an environment characterized by uniform major element, but varying trace element composition. This is not the environment expected in a chamber undergoing simple FC. It is, however, expected in a chamber undergoing continuous fractionation with periodic magma replenishment (24-27). Replenishment supplies the required thermal input; Major elements become "perched" (28) while incompatible trace elements continue to evolve.

PERIODIC REPLENISHMENT MODEL

Detailed microprobe traverses across single plagioclase megacrysts reveal subtle but significant compositional variations, manifest as smooth oscillations of 1-2 AN units about a mean value (29). This suggests periodic rather than continuous replenishment. We have developed a geochemical model of periodic replenishment from the physical model presented below.

Figure 6 represents the physical model of replenishment used in this study. Activity begins with the emplacement of magma of composition Co in a shallow crustal chamber (Fig. 6a). The density of Co is greater than the country rocks and as the magma was emplaced during extensional tectonism, it could not have risen as an independent blob. Rather, it must be supported by a magma column plumbed into a deeper chamber (?) fed from upwelling melting mantle. Although perhaps occurring continuously, we consider the following processes to proceed as steps. Fractional crystallization proceeds in the floor, sidewall and roof region (Fig. 6b). We assume that during this crystallization interval, no leakage occurs (27, 30). We further assume that mixing within the chamber proceeds so as to produce uniform derivative melt (Cl) after (1-F) of crystallization. After this period of crystallization, a replenishing event occurs in which a mass (Mr) of replenishing magma of composition Co enters the chamber (Mr may be $>$, $=$ or $<$ 1-F) and that Co and Cl mix uniformly (31) to produce Cl' (Fig 6c). During the next cycle, Co mixes with Cl' to produce Cl'', etc. Marginal crystallization reduces chamber volume. Therefore, the replenishment event will result in chamber growth, dike emplacement or eruption to the surface (Fig. 6d). For simplicity we assume that the magma thus emplaced will have the composition Cl'. In

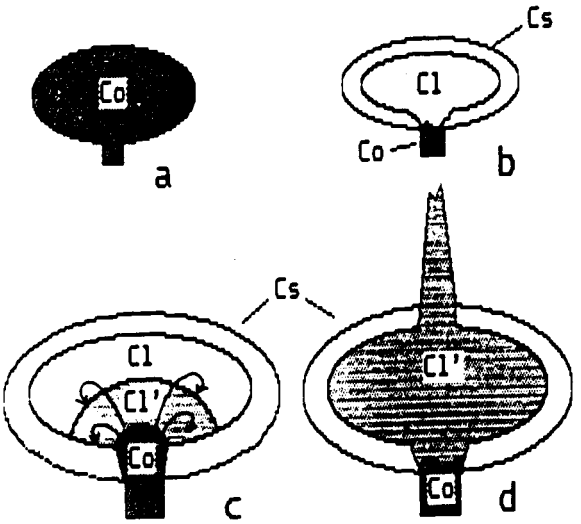


Figure 6. Diagram of replenishment process. Initial magma (Co) is emplaced (a) followed by marginal crystallization (Cs) yielding composition Cl (b). Replenishment by Co with subsequent mixing with Cl yields Cl' (c) which is emplaced as a dike (d). See text.

practice, dike emplacement may occur simultaneously with replenishment, and after various stages of mixing. If mixing is incomplete, either during the fractionation interval, perhaps leading to a stratified magma chamber, or during the replenishment event, compositional endmembers (Co, Cl and Cl') may be forced into the dike conduit at different times (32), perhaps producing the intradike compositional variation observed.

Field evidence supporting multiple injection of magma into the dikes is provided by the non-uniform distribution of megacrysts in the dikes. The megacrysts are often segregated into planar zones subparallel to dike walls. In simplest form, there is only a single planar zone, near the dike center, probably the result of flow differentiation during a single emplacement event. In more complex occurrences, several zones occur at various locations within the dike, and are best explained by periodic pulses of magma entering the dike. As there are no chill zones evident in the complex dikes, the periodicity of emplacement must have been rather frequent. Hunter and Sparks (22) report that zonation in the Hekla (Iceland) system re-established itself in periods of tens to hundreds of years. Although influenced by dike wallrock temperature, if such timing occurred in the MS system, chill zones may not be expected.

Trajectories of periodic replenishment pass through the SEC main group (Figs. 7a,b), consistent with these rocks being derived through this process. Scatter is greater in

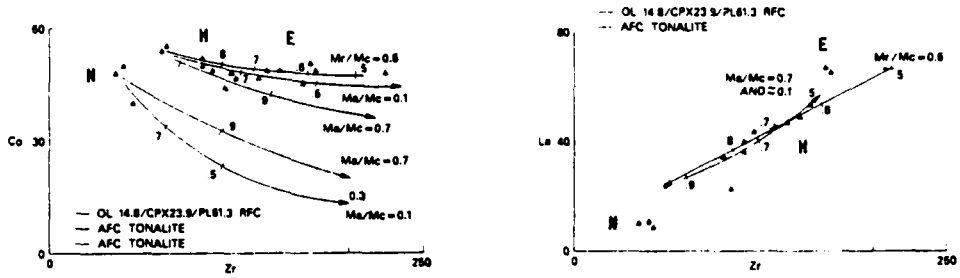


Figure 7. Variation diagrams for the SEC with morb-type, main- and enriched-groups indicated. Modeled are combined replenishment-fractional crystallization (Mr/Mc indicated) and combined assimilation-fractional crystallization (Ma/Mc indicated) using a tonalitic assimilant. Tick marks represent Mm/Mmo values. Fractionating assemblage in the models as indicated.

the Zr-Co plot than in the Zr-La diagram because of the large difference in bulk Kd values for the incompatible-compatible element set. The most compelling geochemical argument for replenishment is the restored agreement between major and trace element models. Above, the decoupling of major and trace elements was discussed in terms of the simple FC model. In particular, F values required by the different data sets did not agree. For the combined replenishment-fractional crystallization model (RFC), however, the F value required by the major elements is within the range of F (actually Mm/Mmo (33)) permitted by the trace elements. These observations, coupled with the requirements of the megacrysts, argue for the RFC model as the controlling process of the SEC main group. The process of periodic replenishment has been recognized in a number of layered mafic intrusions (e.g. 11, 34, 35), primarily as chemical reversals in cumulate sequences (36).

Several studies (e.g. 36, 37) have demonstrated that variations in the parameters of open-system evolution can significantly affect liquid compositions. Modeling during this project concentrated on the impact that varying the mass ratio of replenishing magma to cumulus phases (Mr/Mc) has on concentrations of elements of variable Kds. Modeling indicates that incompatible element ratios are preserved during replenishment. Further, compatible ele-

ments reach steady state concentrations for all mass ratios tested (Mr/Mc = 0.25 - 2.0), although for a given Mr/Mc, more cycles (i.e. a FC interval followed by replenishment) are required to reach steady-state. For a given Mr/Mc, it is the total amount crystallized that is important rather than the number of cycles. In other words, a given mass must be crystallized to reach steady state. If the fractionation interval is 10 percent, it will take twice as many cycles than if the interval is 20 percent. For Mr/Mc > 1.0, even incompatible elements reach steady state although a significantly greater amount of crystallization must occur. The modeling indicates that if one examines variation diagrams where a compatible element (Kd = 3.0) and a moderately incompatible element (Kd = 0.7) are plotted against an incompatible element (Kd = 0.1) the rate of replenishment might be evaluated. Figure 8 compares the predictions of the general model with the actual distribution of Ni (Kd = 2.57), Zn (Kd = 0.5) and Zr (Kd = 0.03) data for the MS dikes. Examination of the model and the

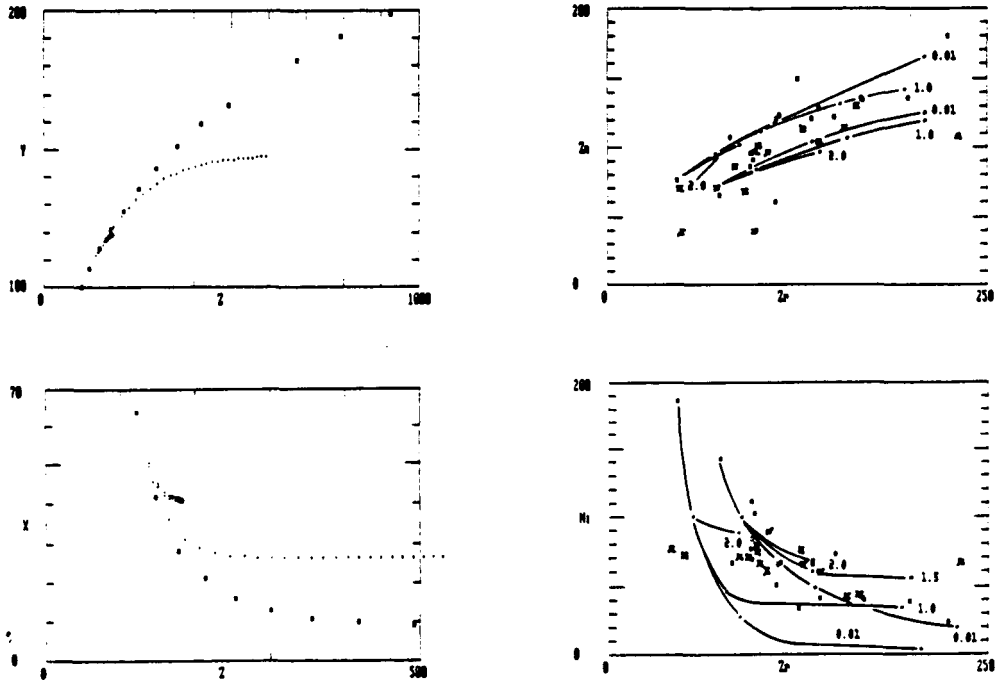


Figure 8. Trajectories of evolution during RFC at Mr/Mc of 0.25 (■), 1.0 (+) and 2.0 (○). Depicted is a compatible element X (Kd = 3.0) and slightly incompatible element Y (Kd = 0.7) plotted against an incompatible element Z (Kd = 0.1). For comparison, Ni (Kd = 2.57) and Zn (Kd = 0.5) against Zr (Kd = 0.03) is shown. See text for further discussion.

data indicate that the dikes may have evolved under variable replenishment rates, although the bulk of the compositions are most consistent with a value of Mr/Mc of less than one.

IMPLICATIONS FOR MULTIPLE PARENT LIQUIDS

The (La/Sm)n versus Zr variation (Fig. 9) reinforces the conclusion that the main group evolved through RFC processes, but clearly indicates that RFC cannot account for compositional differences between the morb-, main- and evolved groups. These intergroup variations may result from source effects, reflecting the trace element characteristics of differing mantle sources tapped during the melting. Alternatively, the compositional differences may reflect crustal interaction between the basalt magma and various crustal components. The possibility that both processes operated is not precluded.

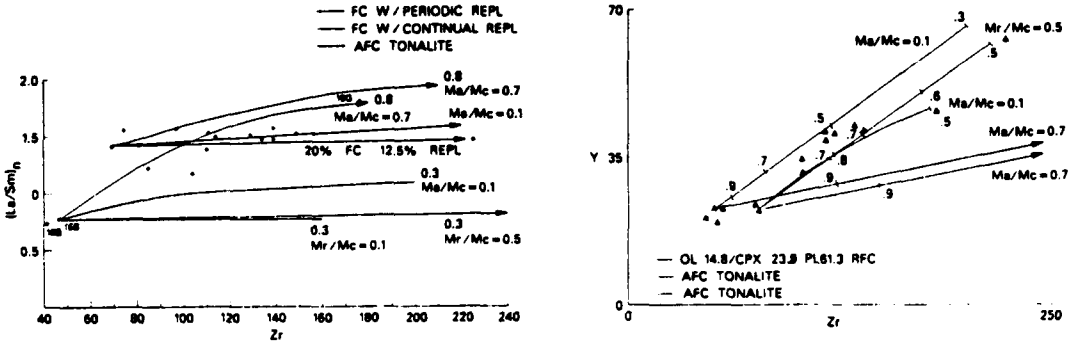


Figure 9. (La/Sm)n against Zr (a) and Y against Zr (b) for the SEC. Also shown are trajectories of various RFC and AFC models (see Fig. 7 and text).

Examination of figures 7 and 9 indicate that only low rates of assimilation (Ma/Mc = 0.1) during combined assimilation-fractional crystallization (AFC) of a tonalitic composition can be accommodated by the main group of the SEC. Higher rates produce unacceptable increases in (La/Sm)n and decreases in Co and Y. This is consistent with the implications of the major element compositions indicative of crystallization in the 3-5 Kb range (see above), where country rock temperatures were probably low. High rates of assimilation (Ma/Mc = 0.7) are capable of relating the morb-type compositions to the main group in

terms of (La/Sm)n and Y, but not for La and Co. High rates of assimilation may be more consistent with emplacement at greater depths (i.e. the crust - mantle boundary) where country rock temperatures are significantly higher. Note that the main group does not lie along the morb trajectories in figure 9, but rather crosses them. This suggests that although the least evolved of the main group may be related to the morb-type compositions through AFC, the remainder of the series is not. The RFC model, with the least evolved of the main group representing Co, remains the most satisfactory model for the SEC main group.

It is apparent from the discussion above, that there is evidence for the existence of more than one "parental" liquid in the SEC, and that these liquids evolved more or less independently. Pertinent to this argument is the observation of the intimate coexistence of both morb-type and the enriched compositions. The REE data for these two groups, shown in figure 2, are from the same dike. No chill zone separates them, and other samples from the same dike have intermediate values. The implication is that the two limiting compositions were contemporaneous and were produced and/or evolved independently. Intermediate compositional groups (the main SEC group and other Ms dikes) might therefore represent variable mixing of the two endmembers. In figure 10, two possible mixing scenarios are displayed. In figure 10a, it is seen that a mixing line between morb-type and enriched compositions define the low-Zr boundary of the SEC data, suggesting that the least evolved, for a given range of Ti/Zr, SEC compositions lie along the mixing line. Also shown is a single FC-controlled trajectory; RFC produces a similar trend. This suggests that the least evolved compositions of the SEC were produced by variable mixing of magmas derived by melts from different sources, one depleted and one enriched. Compositions to the right of the mixing line may have evolved to their respective positions by RFC, related to a specific (?) Co on the mixing line.

As an alternative, figure 10b portrays the mixing of a morb-type melt with a "granitic" crustal component, which could either be a whole rock of this composition, or a crustal melt produced during the underplating of the crust by the basaltic magma. Clearly, the SEC data suggest the potential of crustal input (see also figure 9 and discussion above). FC/RFC trajectories are not shown because these processes have minimal effect on incompatible element ratios. Given the varied character of the Archean crust (e.g. 1, 2 and 7), it might be anticipated that a number of potential mixing lines could be drawn in figure 10b. In this regard, preliminary modeling suggests that mixing of morb-type magmas with amphibolite (Fig. 10b) may

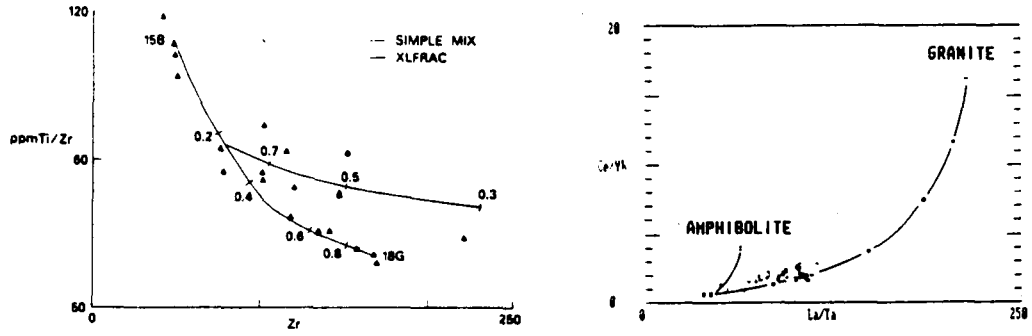


Figure 10. Variation diagrams plotting (Ti/Zr) against Zr, and (Ce/Yb) against (La/Ta) for the SEC. Mixing lines between morb-type (15B) and enriched (18G) compositions (a) and "granite" (b) are shown. Also shown in (a) is the predicted trend of FC (= RFC).

have operated to produce some of the SEC dike compositions.

Distinguishing between the two models portrayed in figures 10a and 10b will probably require isotopic studies. As indicated above, the two models are certainly not mutually exclusive. The compositions of the MS dikes may reflect mixing of multiple endmembers, e.g. depleted mantle, enriched mantle, granitic crust, amphibolitic crust, metasedimentary crust, etc.

CONCLUSIONS

1. The southeast cluster (SEC) of the Matachewan dike swarm can be divided into three compositional groups, based primarily on REE patterns. These groups comprise a morb-type (e.g. light-REE depleted), an enriched group having $(La/Sm)_n > 1.80$, and a main group with intermediate REE patterns.
2. The main SEC group can best be explained by combined replenishment-fractional crystallization (RFC) with variable ratios of replenishment to crystallization; Values were generally less than one. Low assimilation rates are permissible.
3. RFC cannot explain compositional variations between groups of the SEC. Mixing models indicate that mixing of mantle sources and/or interaction of mantle-derived magmas with heterogeneous crustal components may be responsible for intergroup variations.
4. The SEC field and geochemical data suggest a three-

tiered evolution model. The SEC intergroup variation resulted from either magmas being derived from heterogeneous mantle, with at least one component being depleted, or through the interaction of magmas, derived from depleted mantle, with heterogeneous Archean crust. In either of these cases, magma is envisioned to have pooled at the base of the crust. In the former case, the mantle-crust density boundary would trap the various mantle-derived liquids, allowing them to mix in variable proportions, producing the mixing line of figure 10a. In the latter case, the depleted melt would pool and assimilate crustal components, possibly reflected in figure 10b. A combination of both processes is quite possible. Magmas evolve at the base of the crust towards the density minimum. These low-density magmas would accumulate in the upper portions of these subcrustal magma chambers. Additions of new melt from below force these magmas into shallow crustal chambers. Through this re-occurring mechanism, the low-density basalt magmas become the replenishing magmas of the shallow RFC scheme (Fig. 6). Plagioclase megacrysts grow in this "uniform" environment. Dike emplacement, forming the Matachewan swarm, occurs from these shallow chambers. Emplacement occurs as pulses, possibly driven by periodic inputs to the shallow chambers. As a result of evolution and variable mixing within these chambers, successive pulses may differ in composition, leading to intradike variations. Flow differentiation during the periodic pulses of magma emplaced into the dikes produces irregular distribution of phenocryst phases, contributing to the intradike variation.

REFERENCES

1. Ernst, R.E.: Structural and Chemical Studies of Mafic Dike Swarms in Northern Ontario. Ont. Geol. Survey Misc. Pap., No. 106, 1982, pp. 53-56.
2. Phinney, W.C.; Morrison, D.A.; and Maczuga, D.E.: Anorthosites: an Analogue Study. Lun. Sci. Conf. XVIII, 1987, pp. 774-776.
3. Ontario Geological Map - East Central Sheet. Ont. Geol. Surv., Map 2393, 1977.
4. Berger, B.R.: Geology of the Hearst-Kapuskasing Area. Ont. Geol. Surv. Open File Rept., No. 5599, 1986, 88p.
5. Ashwal, L.D.; Morrison, D.A.; Phinney, W.C.; and Wood, J.: Origin of Archean Anorthosites: Evidence from the Bad Vermilion Lake Anorthosite Complex, Ontario. Contrib. Mineral. Petrol., v. 82, 1983, pp. 259-273.
6. Morrison, D.A.; Phinney, W.C.; and Maczuga, D.E.: Archean Anorthosites: Constraints on the Accumulation Process. Lun. Sci. Conf. XVIII, 1987, pp. 670-671.
7. Card, K.D.: Geology and Tectonics of the Archean Superior Province, Canadian Shield. LPI Tech. Rept., No. 86-04, 1986, pp. 27-29.
8. Beakhouse, G.P.; and McNutt, R.H.: Geochemistry of Granitoid Rocks from the Western Superior Province-Evidence for 2- and 3-Stage Crustal Evolution Models. LPI Tech Rept., No. 86-04, 1986, pp. 23-26.
9. Percival, J.A.: Metamorphism and Plutonism in the Quetico Belt, Superior Province, N.W. Ontario. LPI Tech Rept. No. 86-04, 1986, pp. 84-85.
10. Morrison, D.A.; Davis, D.W.; Wooden, J.L.; Bogard, D.D.; Maczuga, D.E.; Phinney, W.C.; and Ashwal, L.D.; Age of the Mulcahy Lake Intrusion, Northwest Ontario, and Implications for the Evolution of Greenstone-Granite Terrains. Earth Planet. Sci. Lett., v. 73, 1985, pp. 306-316
11. Morrison, D.A.; Maczuga, D.E.; Phinney, W.C.; and Ashwal, L.D.: Stratigraphy and Petrology of the Mulcahy Lake Layered Gabbro: An Archean Intrusion in the Wabigoon Subprovince, Ontario. J. Petrol., v. 27, 1986, pp. 303-341.
12. Hodgson, C.J.: The Structure and Geological Development of the Porcupine Camp-A Re-Evaluation. Ont. Geol. Surv. Misc. Pap., No. 110, 1983, pp. 211-225.
13. D.A. Morrison: Personal Communication, 1987.
14. Pearce, J.A.; and Cann, J.R.: Tectonic Setting of Basic Volcanic Rocks Determined Using Trace Element Analyses. Earth Planet. Sci. Lett., v. 19, 1973, pp. 290-300.
15. Grachev, A.F.; and Fedorovsky, V.S.: On the Nature of Greenstone Belts in the Precambrian. Tectonophysics, v. 73, 1981, pp. 195-212.

16. Basaltic Volcanism on the Terrestrial Planets. Pergammon Press, Inc., 1981, 1286 pp.
17. Nelson, D.O.; and Nelson, K.L.: Geochemical Comparison of Alkaline Volcanism in Oceanic and Continental Settings: Clarion Island versus the Eastern Trans-Pecos Magmatic Province. Mantle Metasomatism and Alkaline Magmatism, E.M. Morris and J.D. Pasteris, eds., Geol. Soc. America Spec. Paper 215, in press.
18. Huppert, H.E.; and Sparks, R.S.J.: Restrictions on the Compositions of Mid-Ocean Ridge Basalts: A Fluid Dynamical Investigation. Nature, v. 286, 1980, pp. 46-48,
19. Sparks, R.S.J.; Meyer, P.; and Sigurdsson, H.: Density Variation Amongst Mid-Ocean Ridge Basalts: Implications for Magma Mixing and the Scarcity of Primitive Lavas. Earth Planet. Sci. Lett., v. 46, 1980, pp. 419-430.
20. Fratta, M.; and Shaw, D.M.: 'Residence' Contamination of K, Rb, Li, and Tl in Diabase Dikes. Can. J. Earth Sci., v. 11, 1974, pp. 422-429.
21. Grove, T.L.; and Baker, M.B.: Phase Equilibrium Controls on the Tholeiitic Versus Calc-Alkaline Differentiation Trends. J. Geophys. Res., v. 89, 1984, pp. 3253-3274.
22. Hunter, R.H.; and Sparks, R.S.J.; The Differentiation of the Skaergaard Intrusion. Contrib. Mineral. Petrol., v. 95, 1987, pp. 451-461.
23. Maczuga, D.E.; Morrison, D.A.; and Phinney, W.C.: Rare Earth Elements in Plagioclase Megacrysts and Their Implications for Archean Anorthosite Genesis. Unpub. Ms.
24. Rhodes, J.M.; Dungan, M.A.; Blanchard, D.P.; and Long, P.E.: Magma Mixing at Mid-Ocean Ridges: Evidence from Basalts Drilled near 22 N on the Mid-Atlantic Ridge. Tecto-nophysics, v. 55, 1979, pp. 35-61.
25. Walker, D.; Shibata, T.; DeLong, S.E.: Abyssal Tholeiites from the Oceanographer Fracture Zone. Contrib. Mineral. Petrol., v. 70, 1979, pp. 111-125.
26. O'Hara, M.J.; and Matthews, R.E.: Geochemical Evolution in an Advancing, Periodically Replenished, Periodically Tapped, Continuously Fractionated Magma Chamber. J. Geol. Soc. London, v. 138, 1981, pp. 237-278.
27. Robson, D.; and Cann, J.R.: A Geochemical Model of Mid-Ocean Ridge Magma Chambers. Earth Planet. Sci. Lett., v. 60, 1982, pp.93-104.
28. O'Hara, M.J.: Geochemical Evolution During Fractional Crystallization of a Periodically Refilled Magma Chamber. Nature, v. 266, 1977, pp. 503-507.
29. Maczuga, D.E.; and Morrison, D.A.: Unpublished Data
30. Cann, J.R.: Rayleigh Fractionation with Continuous Removal of Liquid. Earth Planet. Sci. Lett., v. 60, 1982, pp. 114-116.

31. Campbell, I.H.; and Turner, J.S.: Turbulent Mixing Between Fluids with Different Viscosities. Nature, v. 313, 1985, pp. 39-42.
32. Blake, S.; and Campbell, I.H.: The Dynamics of Magma-Mixing During Flow in Volcanic Conduits. Contrib. Mineral. Petrol., v. 94, 1986, pp. 72-81.
33. DePaolo, D.J.: Trace Element and Isotopic Effects of Combined Wallrock Assimilation and Fractional Crystallization. Earth Planet. Sci. Lett., v. 53, 1981, pp. 189-202.
34. Pallister, J.S.; and Hopson, C.A.: Samail Ophiolite Plutonic Suite: Field Relations, Phase Variations, Cryptic Variation and Layering, and a Model of a Spreading Ridge Magma Chamber. J. Geophys. Res., v. 86, 1981, pp. 2593-2644.
35. Elthon, D.; Casey, J.F.; and Komor, S.C.: Cryptic Mineral Chemistry Variations in a Detailed Traverse Through the Cumulate Ultramafic Rocks of the North Arm Mountain Massif of the Bay of Islands Ophiolite, Newfoundland. Ophiolites and Oceanic Lithosphere, I.G. Gass, S.J. Lippard and A.W. Shelton, eds., Oxford, England, Blackwell, 1984, pp. 83-100.
36. Thy, P: Magmas and Magma Chamber Evolution, Troodos Ophiolite, Cyprus. Geology, v. 15, 1987, pp. 316-319.
37. Karson, J.A.; and Elthon, D.: Evidence for Variations in Magma Production Along Oceanic Spreading Centers: A Critical Appraisal. Geology, v. 15, 1987, pp. 127-131.

MID-TERTIARY VOLCANO-TECTONIC DEVELOPMENT OF THE SOUTHWESTERN CORDILLERA OF NORTH AMERICA

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Kerri L. Nelson, MSc.
Academic Rank:	Instructor
University & Department:	Sul Ross State University Analytical Laboratory Alpine, Texas 79832
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
JSC Colleague:	Charles A. Wood, PhD.
Date:	August 21, 1987
Contract Number:	NGT 44-001-800

Abstract

In the Southwestern Cordillera (SC) of North America, volcanic style changed from dominantly calcalkaline stratovolcanoes to caldera-related magmatism during the mid-Tertiary. The dominant tectonic process affecting the region during this time was convergence of the Farallon and North American Plates. The calderas are spatially overprinted on the earlier calcalkaline volcanics and, in some instances, there is a temporal overlap between the two contrasting styles of volcanism. The change in style of volcanism indicates a change in the operative stress regime: compressional for the earlier calcalkaline volcanism and tensional for development of the calderas.

The major volcanic centers in the region are the Sierra Madre Occidental (Mexico), Trans-Pecos Province (Texas), Mogollon-Datil field (New Mexico), San Juan Volcanic Field (Colorado) and the Marysvale Volcanic Field (Utah). We have compared the development of these centers to 1) evaluate the volcano-tectonic relationship of caldera development within and between centers and 2) determine the relationships between the earlier calcalkaline and later caldera-style volcanism.

The calderas exhibit three distinct stages of development that are closely associated with the East Pacific Rise/trench collision. Calderas formed prior to impact tended to be small both in number and size. The largest and most numerous calderas developed close to impact of the rise. After impact of the rise, an abrupt decrease in caldera development occurred. Variation in convergence rates between the Farallon and North American Plates significantly influenced the style of volcanism. Extremely rapid convergence rates (>15 cm/yr) resulted in compressive stress and formation of stratovolcanoes. At ~40 ma, convergence rates slowed, causing an increase in dip of the Benioff and establishment of an extensional regime. The spatial and temporal association of the calcalkaline and caldera-related volcanism argues for the SC representing a region of continued arc magmatism in which the style of volcanism varied in response to differences in regional stresses.

Introduction

During the mid-Tertiary, there was extensive caldera-related volcanism in the Southwestern Cordillera (SC) of North America. The primary caldera centers in this region are the Marysvale, San Juan, Mogollon-Datil, Trans-Pecos and Sierra Madre Occidental volcanic fields. In excess of 3 million cubic km of ignimbrites were erupted from these centers during the Oligocene-Early Miocene.

Although this region is the largest terrestrial caldera field known and has been intensively studied, there is no general agreement upon the tectonic environment of caldera formation. Some authors favor caldera development in a back-arc region [1,2] while others believe that the calderas formed in the area of the primary arc [3] or intra-arc [4]. We have examined the volcano-tectonic history of the various centers in the SC in order to 1) better understand the relationship between the calderas and subduction processes and 2) better delineate the tectonic setting of both the earlier calcalkaline and later caldera-related volcanism.

Regional Setting

The SC is defined here as the area from the Sierra Madre Occidental in western Mexico through and including the Marysvale Volcanic field of Utah (Fig. 1). The major centers in this region are the Sierra Madre Occidental (SMO), Trans-Pecos Province (TP), Mogollon-Datil (MD), San Juan Volcanic field (SJ) and Marysvale Volcanic field (MV). These centers were chosen as they 1) mark the eastern boundary of Tertiary-age magmatism in the SC, 2) have associated caldera development and 3) represent distinct geographical provinces, thus facilitating comparison of regional variations. Data used to characterize these fields was acquired from a number of sources [5-14] which greatly aided our regional synthesis. A comparison of pertinent features of each of the centers is given in table 1.

The SC of North America was volcanically active from approximately 80 ma to the present [15]. During this time however, volcanism changed from subduction-related to extensional (Basin and Range and Rio Grande Rift). The precise timing of this transition is uncertain, but inception of Basin and Range tectonism at 32 to 30 ma in the southern portion of this region [16-18] followed by a northward migration of extension related activity is generally agreed upon.

Volcanic activity can be separated into three main periods. Early magmatism in the region was primarily

Figure 1. Distribution of Caldera Fields in the Southwestern Cordillera

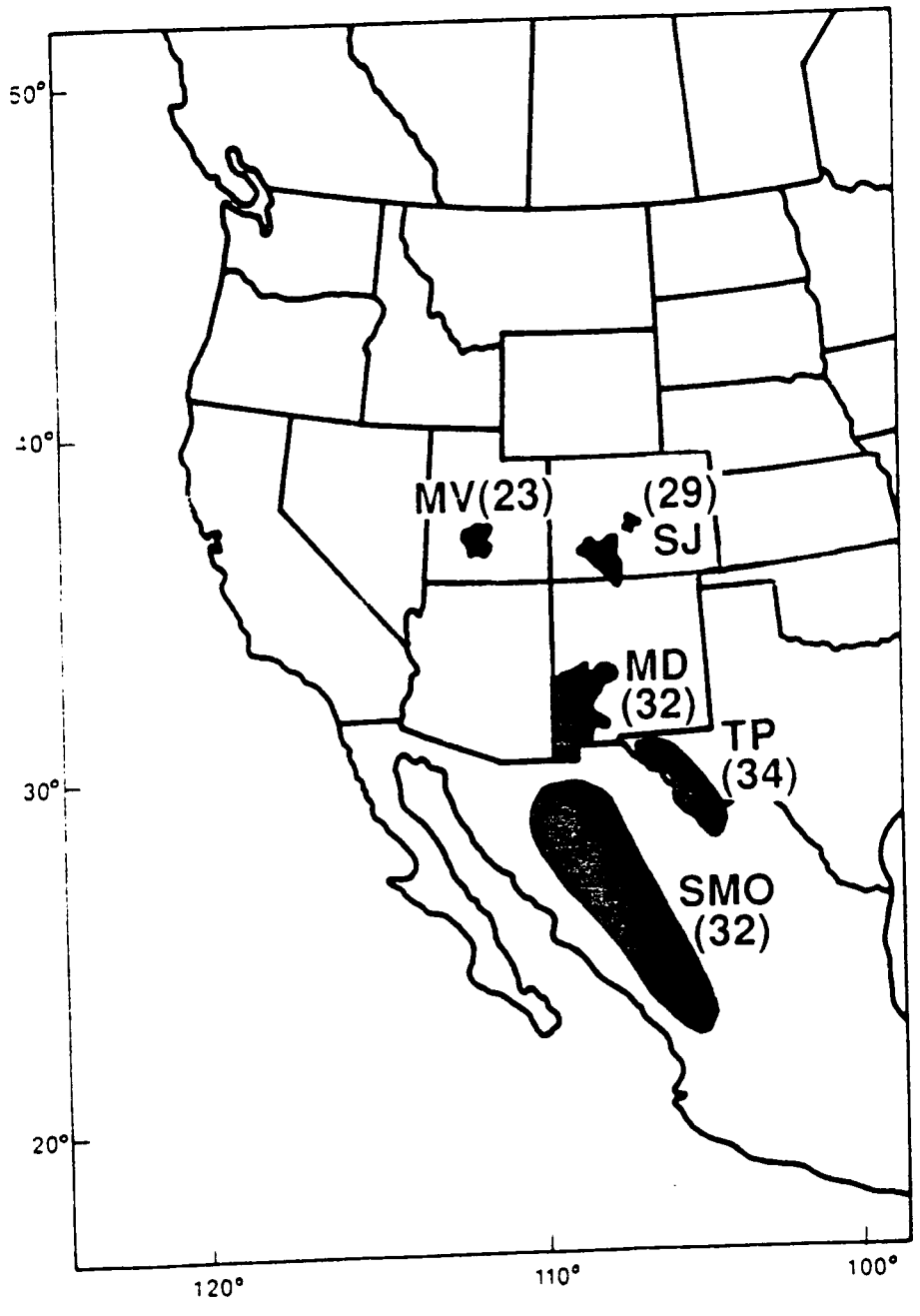


Table One. Comparison of Caldera Centers in the Southwestern Cordillera.

Center	Location	Age[1]	Caldera[2]	Batholith	Earlier Volcanism
Marysvale	Utah	23	6	Yes	Andesitic Stratovolcanoes
San Juan	Colorado	29	16	Yes	Andesitic Stratovolcanoes
Mogollon-Datil	New Mexico	32	28	Yes	Andesitic Stratovolcanoes
Trans-Pecos	Texas	34	12	No	Basaltic Flows
Sierra Madre Occidental	Mexico	32	12/300	Yes	Andesitic

1 Average age of caldera field based on reported age dates.

2 These numbers represent minimum values. More detailed mapping in each field may facilitate recognition of more calderas. For instance, in the Sierra Madre Occidental, 12 calderas have been located although in excess of 300 are thought to be present.

References: Swanson et al., 1978; Cameron et al., 1980; Elston, 1984; Henry and Price, 1984; Lipman, 1984; Steven et al., 1984; Swanson and McDowell, 1984; Varga and Smith, 1984.

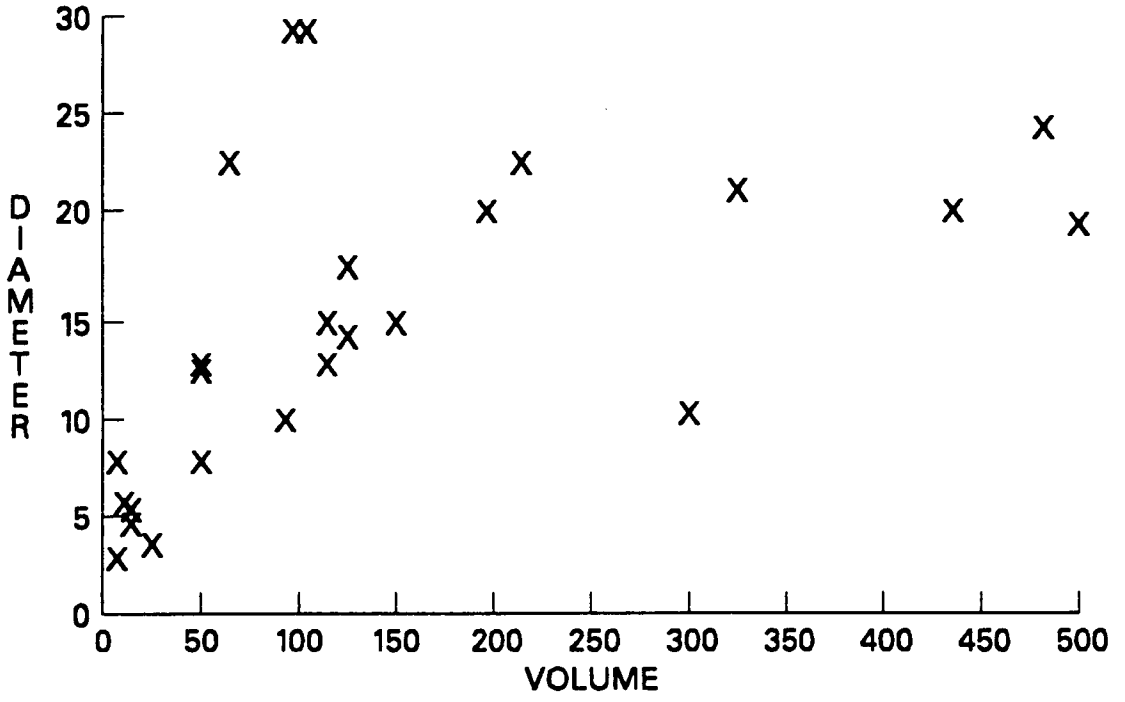
calcalkaline in composition, dominated by andesitic to dacitic stratovolcanoes, cones and domes. Widespread caldera development followed the main period of calcalkaline activity, although there was some overlap between the two phases of magmatism in some centers [1,12].

Ignimbrites erupted during this time varied compositionally from alkaline to calcalkaline in composition. The final period of volcanism is characterized as bimodal. Compositionally, basalts, basaltic andesites and high-silica rhyolites are the dominant eruptive products [e.g. 19]. The two initial stages of volcanism are believed to be related to subduction processes, while the later stage is related to extensional tectonics [4,20].

The caldera activity was transitional between the early arc volcanism and later extensional Basin and Range/Rio Grande Rift activity. Several individuals have noted that a "magmatic lull" existed between the stage of caldera development and extensional volcanism and that the calderas are cut by Basin and Range faulting [1,3]. Therefore, we regard the development of calderas in the SC as significant in the timing of transition of volcanic and tectonic activity in the region.

To better evaluate the volcano-tectonic development of the SC, we have taken a regional approach and gathered data from over 75 calderas in the major caldera centers. Data on caldera age, diameter, volume, distance from trench, latitude, longitude, previous volcanic activity and elemental and isotopic composition were compiled There is not a complete data set available for many of the calderas, which limits the degree of intra and intercenter comparisons. For a significant number or calderas, there is a lack of data regarding the volume of eruptive products. There are three main reasons for the paucity of data for this parameter. First, significant erosion has occurred since emplacement and volume calculations are minimum estimates. Secondly, in some areas, field mapping is not complete and the actual distribution of ignimbrite sheets is unknown. Finally, several calderas in the SMO have been identified exclusively through remote sensing techniques. As a result, we have volume data on less than half of the calderas in the SC centers. Volume, is an important parameter as the volume of erupted products approximates the size of the magma chamber beneath the caldera. For our data set, a positive correlation between diameter and volume for calderas up to 20 km in diameter exists (Fig. 2). For calderas above this size, small changes in caldera diameter can result in very large changes in erupted volume. The correlation between

Figure 2. Relationship between caldera volume and diameter.



diameter and volume is not unique to the calderas we evaluated as other workers have noted similar relationships [21,22]. To work with a complete data set, we will use diameter as an approximation of volume and consider those two parameters jointly in the remaining discussion.

Tectonic Setting

The primary tectonic control on volcanism in the SC between 80 and 30 ma was convergence between the Farallon and North American Plates [e.g. 23]. Convergence rates between these two plates are believed to have been quite rapid (>15 cm/yr) between 60 and 40 ma [24,25] which caused volcanism to migrate a significant distance landward as the Benioff dip shallowed [26]. At approximately 40 ma, subduction slowed to less than 10 cm/yr in response to the Farallon Plate being younger, hotter and therefore, more difficult to subduct. This resulted in a steepening of the Benioff. The decrease in convergence rate continued until the East Pacific Rise (EPR) intersected the trench causing a cessation of subduction along that portion of the continental margin. This places tight constraints on the timing of volcanism and the volcano-tectonic relationships. The majority of the primary arc volcanoes were emplaced during periods of rapid convergence, while the majority of the calderas developed after convergence rates had slowed. This suggests a very intimate relationship between development of caldera volcanism and convergence rates existed.

Discussion

At subduction zones, a continuum between compressional and extensional tectonics exists. A possible mechanism for creating the shift in stress regime is a decrease in orogenic stresses resulting from the slowing of convergence rates [27]. Initially fast rates of convergence and shallow dip angles favors establishment of volcanism landward. Once convergence slows and the dip angle increases, extensional stress regimes can be established. This would lead to crustal thinning on the continent. The process of subduction hinge migration outlined above is generally cited as an important mechanism for opening of the back-arc [e.g. 28] but is evidently applicable in the SC.

Crustal thinning in response to decreased convergence rates would enhance the emplacement of magma chambers at shallow depths. The development of extensional regimes

allows more mantle magma access to the continental crust, promoting subcrustal heating of the lower crust which may lead to the development of initial abundant silicic volcanics [22]. Further, modeling of lithospheric magmatism by other authors support the mechanism of crustal thinning as important in the formation of caldera development [29].

If the relationship between crustal thinning and slowing of subduction is related to caldera formation, there should be a correlation between caldera age and development of extensional stress regimes. To evaluate this, we have defined a parameter termed here as normalized age, $(AGE)_n$. This parameter normalizes the age of a given caldera to the timing of the EPR/trench collision. An age of 30 ma was chosen for the collision based on models of plate reconstructions [20,23]. The age of 30 ma is not precise as intersection of the EPR varied both temporally and spatially, but is a reasonable approximation for the majority of the SC, considering processes at the trench will not be felt instantaneously on the continent. Values of $(AGE)_n$ >1 indicate a caldera formed prior to intersection of the EPR, while values <1 indicate caldera formation postdates the collision. When $(AGE)_n$ is plotted against caldera diameter for calderas in the SC, three consistent stages of caldera development are observed (Fig. 3). The earliest calderas (stage 1) formed in the SC tend to have small diameters (<15 km). Caldera diameters reach their greatest size (>25 km) during stage 2 when $(AGE)_n$ is 1.1 to 0.9. Stage 2 also correlates with the most intense period of caldera development/volcanism. Caldera development decreased abruptly during stage 3. The relationships outlined above are not an artifact of scatter from plotting several different centers as the trends are characteristic of individual centers as well (Fig. 4 a,b).

We believe that the data support the model outlined previously and illustrated in figure 5. Stage 0 reflects rapid convergence rates and compressional tectonics, represented by the formation of dominantly calcalkaline stratovolcanoes and cones. As subduction slowed (stage 1) a relaxation of compressive stresses at the trench caused development of neutral to mildly extensional stresses inland. Under these conditions, small magma chambers were emplaced at shallow levels in the crust, hence small calderas were developed. Stage 2 represents the continual slowing of subduction until impact of the EPR at the trench. This would correlate with further development of extensional regimes (increased crustal thinning) and the shallow emplacement of larger magma chambers, perhaps of

Figure 3. Covariation of (AGE)n and diameter for all calderas in the SC caldera centers.

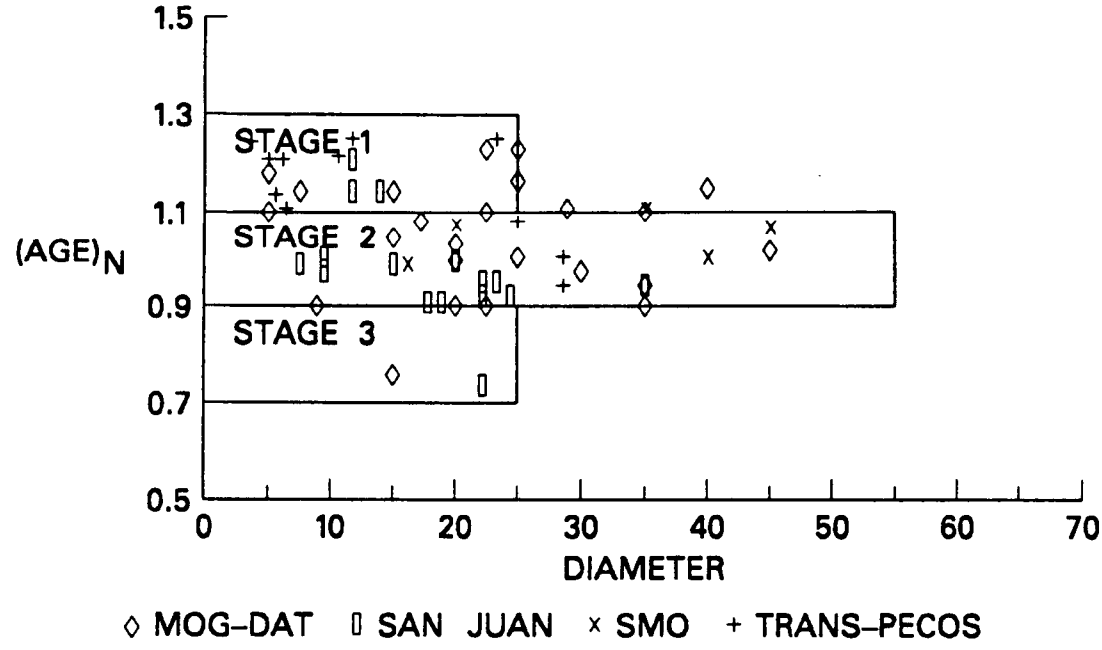
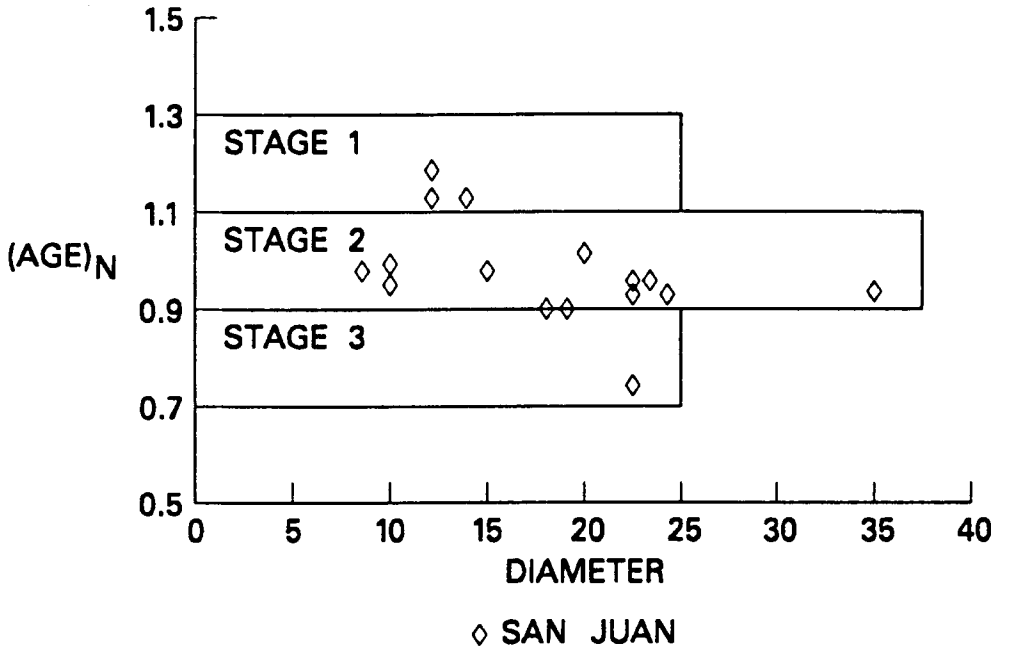


Figure 4. (AGE)n vs. diameter for a) the San Juan Calderas and b) the Mogollon-Datil calderas.

a.



b.

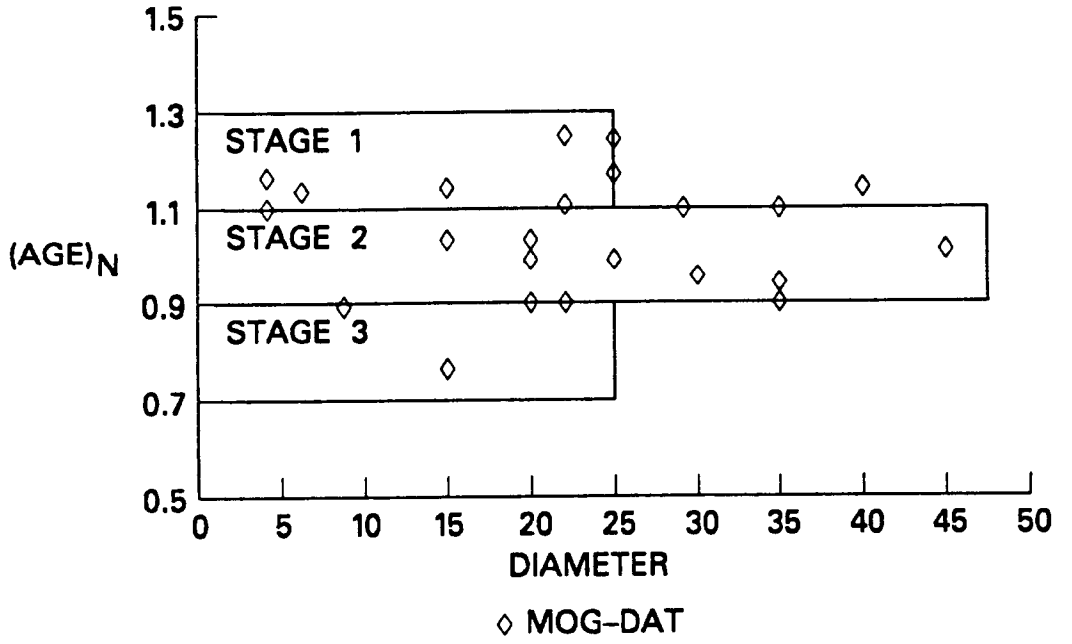
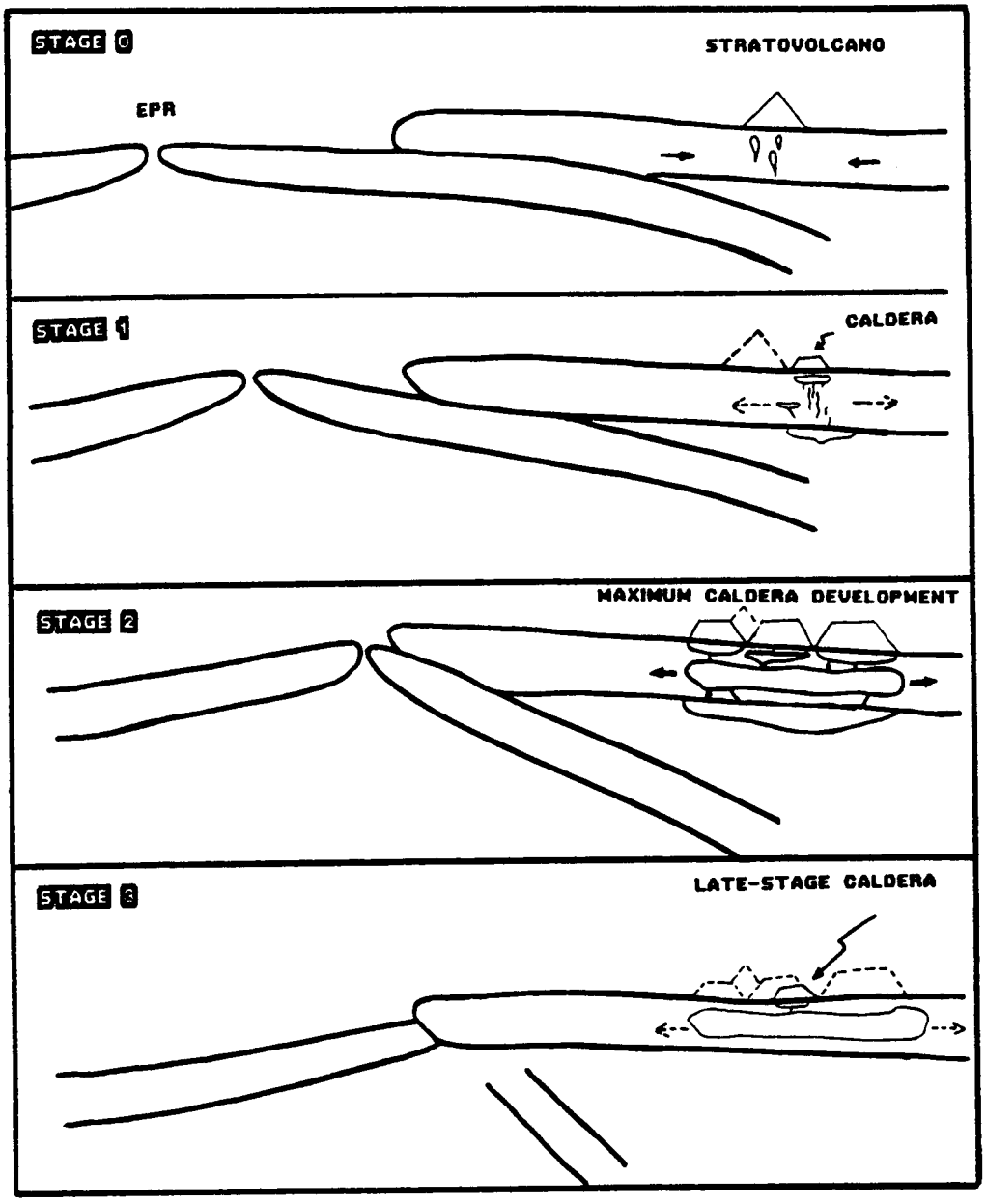


Figure 5. Variations in volcanic development in response to variable stress regimes in the primary arc. Explanation in text.



batholithic proportions. After impact of the EPR at the trench, subduction along that portion of the continental margin ceased (stage 3). One might expect that the transition from late stage 2 to stage 3 would be marked by development of smaller calderas. There are very few data points falling on or below the boundary between of stage 2 and 3. Those that do fall in this range are, for the most part, less than 25 km in diameter. Some of these stage 3 calderas may be young enough to be related to later extensional tectonics (e.g. Lake City Caldera, Ref. 10). However, others falling in this group are cut by later faults, indicating they developed prior to Basin and Range or Rio Grande Rift processes.

Two mechanisms, possibly interrelated, may be responsible for the decrease in late-stage caldera development. First, as subduction initially slowed, then ceased, the role of the subducted slab in the generation of magma would have been significantly reduced (refer to Fig. 5). A decrease in magma generation could account for the presence of smaller magma chambers and a lessening of volcanic activity. Alternatively, the extensional stress associated with the slowing of subduction would not be maintained once subduction had stopped. In this scenario, even if large magma bodies were present at depth, they would not have been able to easily ascend if extensional stresses were not great.

Several authors have noted that there is a relationship between the calderas and shallowly emplaced batholiths in the SC [e.g. 8,10]. These batholiths are believed to represent large-scale chambers which supplied magma to the calderas. Importantly, late-stage calderas are dominantly found within clusters of older calderas, indicating that the younger calderas are also related to the batholiths and may have exploited previously established magmatic conduits. The abrupt cessation of volcanism and the relationships between batholiths and calderas argue for the stress regime exerting the major control on caldera development. The presence of the batholiths indicates magmas must have been present throughout the history of the caldera field, but were most successfully tapped during stages of extensional stress.

As a final point, the presence of both compressional and extensional tectonics and magmatism in the SC has been a point of controversy for some time. It has been argued that the SC represents either a primary arc [e.g. 3], back arc [e.g. 1] or intra-arc extension [e.g. 4]. We believe that the distinction is basically one of semantics.

As distance from the trench remained constant, it is to consider an area initially an arc and then a back arc based solely on the change in volcanic style resulting from differences in stress regime. Further, if the calderas represent a back-arc assemblage, there is no complimentary "primary arc" positioned trenchward. One problem in accepting the SC volcanic centers as primary arc components has been the great distance (up to 1000 km) that both the calcalkaline volcanoes and calderas are found from the trench. However, much of this distance may have been acquired during or post emplacement of the SC volcanics through processes of accreted terranes and pre- and post-Basin and Range extension [2,4,30,31]. If the combined effect of these processes are accounted for, the position of the SC would be much closer to the trench. The temporal and spatial association of calcalkaline volcanism to the calderas is more consistent with the entire volcanic suite representing the primary arc. The term intra-arc spreading seems most appropriate as it implies that the primary site of arc volcanism remained fixed while the change in volcanic style resulted primarily from a change in tectonic stresses (compressional to extensional).

The only center that does not fit the general scheme outlined above is the TP province. It is anomalous in that it lacks a well defined calcalkaline assemblage and has no associated batholith [8]. Also, the average age of the TP calderas is older than any of the other centers. Based on both chemical and geographical trends, it has been suggested that the TP is closely associated with the SMO field [e.g. 32-34] and may represent a back arc assemblage. As stress orientations indicate that the bulk of the TP calderas were emplaced under compressive stresses and it also has been suggested that the TP are part of the primary continental arc [3]. Under conditions of large convergence rates and a low Benioff dip angle, compression will occur in the back-arc [35]. Since the SMO are trenchward from the TP and chemical gradations between the two provinces exist, we believe that the TP calderas were formed in the back-arc environment and that the SMO-TP represent one large volcanic province.

Conclusions

The volcano-tectonic development of the SC reflects the change in convergence rates between the Farallon and North American Plates. Rapid convergence rates and compressional stresses favored development of a calcalkaline magmatic arc. As subduction slowed in response to subduction of progressively hotter and younger portions of the Farallon Plate, an extensional stress regime was established in the arc area and caldera development was initiated.

The variation in caldera size and activity is related to the amount of extension that was experienced in the arc. The timing of the EPR/trench collision apparently marked the maximum convergence-related extension felt on the continent and is associated with the most numerous and voluminous period of caldera development.

The temporal and spatial relationships between the calcalkaline volcanoes and calderas indicates that the site of magmatism remained fixed. Therefore, the changing volcanic styles do not represent arc vs. back-arc assemblages. Rather, they reflect volcanic development in an arc environment that varied in response to changes in regional stresses.

References Cited

1. Elston, W.E.; and Bornhorst, T.J.: The Rio Grande Rift in Context of Regional Post-40 m.y. Volcanic and Tectonic Events. Rio Grande Rift: Tectonics and Magmatism, Reicker, R.E., ed., Am. Geophys. Union, Washington D.C., 1979, pp. 416-438.

2. Elston, W.E.: Subduction of Young Oceanic Lithosphere and Extensional Orogeny in Southwestern North America during Mid-Tertiary Time. Tectonics, vol. 3, 1984. pp. 229-250.

3. Price, J.G.; and Henry, C.D.: Stress Orientations during Oligocene volcanism in Trans-Pecos Texas: Timing the transition from Laramide compression to Basin and Range tension. Geology, vol. 12, 1984, pp. 238-241.

4. Zoback, M.L.; Anderson R.E.;and Thompson, G.A.: Cainozoic evolution of the state of stress and style of tectonism of the Basin and Range province of the western United States. Phil. Trans. R. Soc. Lond., vol. A 300, 1981, pp. 407-434.

5. Swanson, E.R.; Keizer, R.P.; Lyons,J.I.; and Clabaugh, S.E.:Tertiary volcanism and caldera development in the Durango City area, Sierra Madre Occidental, Mexico. Geol. Soc. Am. Bull., vol. 89, 1978, pp. 1000-1012.

6. Cameron, K.L.; Cameron, M.; Bagby, W.C.; Moll, E.J.; and R.E. Drake: Petrologic characteristics of mid-Tertiary volcanic suites, Chihuahua, Mexico. Geology, vol. 8, 1980, pp. 87-91.

7. Elston, W.E.: Mid-Tertiary age Ash Flow Tuff Cauldrons, Southwestern New Mexico. J. of Geophys. Res., vol. 89 1984, pp. 8733-8750.

8. Henry, C.D.; and Price, J.G.: Variations in Caldera Development in the Tertiary Volcanic Field of Trans-Pecos Texas. J. Geophys. Res., vol. 89, 1984, pp. 8765-8786.

9. Cepeda, J.: Geology and Geochemistry of the Igneous Rocks of the Chinati Mountains, Presidio County, Texas: PhD. Dissertation, Univ. Texas at Austin, 1977.

C-2

10. Lipman, P.W.: The Roots of Ash Flow Calderas in Western North America: Windows Into the Tops of Granitic Batholiths. J. Geophys. Res.,vol. 89, 1984, pp. 8801-8841.

11. Steven, T.A.; Rowley, P.D.; and Cunningham, C.G.: Calderas of the Marysvale Volcanic Field, west Central Utah, J. Geophys. Res.,vol. 89, 8751-8764, 1984.

12. Swanson, E.R. and McDowell, F.W.:Calderas of the Sierra Madre Occidental Volcanic field, western Mexico. J. Geophys. Res., vol. 89, 1984, pp. 8787-8799.

13. Varga, R.J.; and Smith, B.M.: Evolution of the Early Oligocene Bonanza Caldera, Northeast San Juan Volcanic Field, Colorado. J. Geophys. Res., vol. 89, 1984, pp. 8679-8694.

14. Lindsey, D.A.: Tertiary volcanic rocks and uranium in the Thomas Range and northern Drum Mountains, Juab County, Utah. U.S. Geol. Surv. Prof. Pap. 1221, 1982.

15. Lipman, P.W.; Prostka, H.J.: and Christiansen, R.L.: Cenozoic volcanism and plate-tectonic evolution of the western United States. I. Early and middle Cenozoic. Philos. Trans. Roy. Soc. London, vol. A 271, 1972, pp. 217-248.

16. Chapin, C.E.: Evolution of the Rio Grande Rift-A Summary. Rio Grande Rift: Tectonics and Magmatism, Reicker, R.E., ed., Amer. Geophys. Union, Washington, D.C., 1979, pp. 1-5.

17. Laughlin, A.W.; Aldrich, M.J.;and Vaniman, D.: Tectonic implications of mid-Tertiary dikes in west-central New Mexico. Geology, vol. 11, 1983, pp. 45-48.

18. Eaton, G.P.: The Miocene Great Basin of Western North America as an extending Back-arc region. Geodynamics of Back arc Regions, Carlson R.L. and Kobayashi, K., eds., Tectonophys., vol. 102,1984, pp. 275-296.

19. Noble, D.C.: Some observations on the Cenozoic Volcano-Tectonic evolution of the Great Basin, Western United States. Earth Planet. Sci. Lett., vol. 17, 1972, pp. 142-150,

20. Lipman, P.W.: Cenozoic Volcanism in the Western United States: Implications for Continental Tectonics, Continental Tectonics, Burchfiel, B.C. Oliver, J. and Silver, L.T. co-chairs, Natl. Acad. Sci., Washington, D.C., 1980, pp. 161-174.

21. Smith, R.L.: Ash Flow Magmatism. Ash Flow Tuffs, Chapin, C.E. and Elston, W.E., eds, Geol. Soc. Am. Spec. Paper 180, 1979, pp. 5-27.

22. Cas, R.A.F.; and Wright, J.V.: Volcanic Successions. Allen and Unwin, London, 1987.

23. Atwater, T.: Implications of plate tectonics for the Cenozoic evolution of western North America. Geol. Soc. Am. Bull., vol. 81, 1970, pp. 3513-3536.

24. Carlson, R.L.: Cenozoic convergence along the California coast: a qualitative test of the hot-spot approximation. Geology, vol. 10, 1982, pp. 191-196.

25. Jurdy, D.M.: The Subduction of the Farallon Plate beneath North America as derived from Relative Plate Motions. Tectonics, vol. 3, 1984, pp. 107-113.

26. Coney, P.J.; and Reynolds, S.J,: Cordilleran Benioff zones. Nature, vol. 270, 1977, pp.403-406.

27. Dickinson, W.R.: Plate Tectonic Evolution of the Southern Cordillera. Relations off Tectonics to Ore Deposits in the Southern Cordiller, Dickinson, W.R. and Payne, W.D., eds., Ariz. Geol. Soc. Digest, vol. XIV, 1981, pp. 113-136.

28. Carlson,R.L.;and Mella, P.J.:Subduction hinge Migration. Tectonophys., vol. 102, 1984, pp. 399-411.

29. Hildreth, W.: Gradients in silicic magma chambers: Implications for lithospheric magmatism. J. Geophys. Res., vol. 86, 1982, pp. 10153-10192.

30. Coney, P.J.: Circum-Pacific Tectogenesis in the North American Cordillera. Circum-Pacific Orogenic Belts and Evolution of the Pacific Ocean Basin, Monger,J.W. and Francheteau, J., eds., Am. Geophys. Union, Washington, D.C., 1987, pp. 59-69

31. Coney, P.J.; and Harms, T.A.: Cordilleran metamorphic core complexes: Cenozoic extensional relics of Mesozoic compression. Geology, vol. 12, 1984, pp. 550-554.

32. Cameron, K.L; and Cameron, M.: Geochemistry of quartz-normative igneous rocks from the Chinati Mountains and Terlingua areas, west Texas: A comparison with Cenozoic volcanic rocks from Chihuahua and Baja California Sur, Mexico. Igneous Geology of Trans-Pecos Texas: Field Trip Guidebook and Research Articles, Price, J.G., Henry, C.D., Parker, D.F. and Barker, D.S., eds., Texas Bureau of Economic Geology, Austin, Texas, 1986, pp. 143-163.

33. McDowell, F.W.; and Clabaugh, S.E.: Ignimbrites of the Sierra Madre Occidental and their relation to the tectonic history of western Mexico. Ash Flow Tuffs, Chapin, C.E. and Elston, W.E., eds., Geol. Soc. Am. Spec. Paper 180, 1979, p. 113-124.

34. Nelson, D.O.; Nelson, K.L.; Reeves, K.D.; and Mattison, G.D.: Geochemistry of Tertiary alkaline rocks of the eastern Trans-Pecos Magmatic Province, Texas. Contrib. Mineral. Petrol., in press.

35. Shimozuru, D.; and Kubo, N: Volcano Spacing and Subduction. Arc Volcanism: Physics and Tectonics, Shimozuru, D. and Yokoyama, I., eds. Terra Scientific Publishing Company, Tokoyo, 1983, pp. 141-151.

Fault-Tolerant Computer Architecture Based on INMOS Transputer Processor

Final Report

NASA/ASEE Summer Faculty Fellowship Program --1987

Johnson Space Center

Prepared by:	Jorge L. Ortiz, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Electrical and Computer Engineering Department University of Puerto Rico Mayaguez,P. R. 00709-5000
NASA/JSC	
Directorate:	Engineering and Development
Division:	Avionic Systems
Branch:	Flight Data Systems
JSC Colleague:	Michael M. Thomas
Date:	August 7, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

Redundant processing has been used for several years in mission flight systems. In these systems, more than one processor performs the same task at the same time but only one processor is actually in real use. A fault-tolerant computer architecture based on the unique features provided by INMOS Transputers is presented in this report. The Transputer architecture provides several communication links that allow data and command communication with other Transputers without the use of a bus. Additionally the Transputer allows the use of parallel processing to increase the system speed considerably.

The processor architecture consists of three processors working in parallel keeping all the processors at the same operational level but only one processor is in real control of the process. The design allows each Transputer to perform a test to the other two Transputers and report the operating condition of the neighbor processors. A graphic display has been developed to facilitate the identification of any problem by the user.

I. Introduction

The concept of redundant processing has been used in the space for long time specially for critical maneuvers like landing or launching. In these cases, several processors had been working in parallel performing the same task but only one of then is in real control of the process. If something goes wrong with this computer the system operator or astronaut can switch the operation to another processor.

Recently, the C. S. Draper Laboratory designed a fault-tolerant processor called Advance Information Processing System (AIPS) with the concept of maintaining three processors (or more) working redundantly and testing each other to "vote" on the status of the other processors. In this fashion, the user has the information about the system performance on real time. This system is linked by a data communication bus called the Inter-Computer Bus (IC) for communication between processors and other I/O devices.

A fault-tolerant computer architecture based on the unique features provided by INMOS Transputer has shown to be an adequate alternative to this kind of processor. Among the characteristics that can improve the design of the processor are the serial communication links that allow data and command communications with other Transputers without the use of a bus, and the capability of parallel processing to increase the system speed. Therefore, a Transputer Fault-Tolerant Processor (TFTP) designed based on the Transputers could mean a faster more reliable processor.

Discussion and Results

The first objective of this research was to design a fault-tolerant processor with a parallel architecture based on the INMOS Transputer. Two solutions to this problem were presented at the moment. The first one, presented by Mr. Dennis Taylor, uses four T414 transputers with three of then working in parallel an the remaining one will be the coordinator as shown in Figure 1. Three parallel processors will perform the same task while the coordinator will compare the results of the operations and report to the user if it finds a fault in one of the processors. This architecture is very efficient and is easy to keep control on the parallel processors but the possibility exists that the fault could come from the coordinator itself. The second architecture, shown in Figure 2, consist of keeping three processors (or more) working in parallel. All of then will be kept in the same level, but only one of then will perform the real operation. Each of then will keep performing tests to the other two processors and reporting to the operator the results of these tests. The three Transputers are interfaced to the Transputer Development System (TDS) in an IBM AT compatible. The design allows each Transputer to perform an evaluation to the other two Transputers and report the operating condition (based on that test) of the neighbor Transputers. The test consists of sending an integer constant to a processor and the processor under test will return its square value. This result is analyzed and compared with the previously known solution to later send a report to the host computer. At this moment, three Transputers are running in parallel, performing the indicated test, and finally showing its report on the screen.

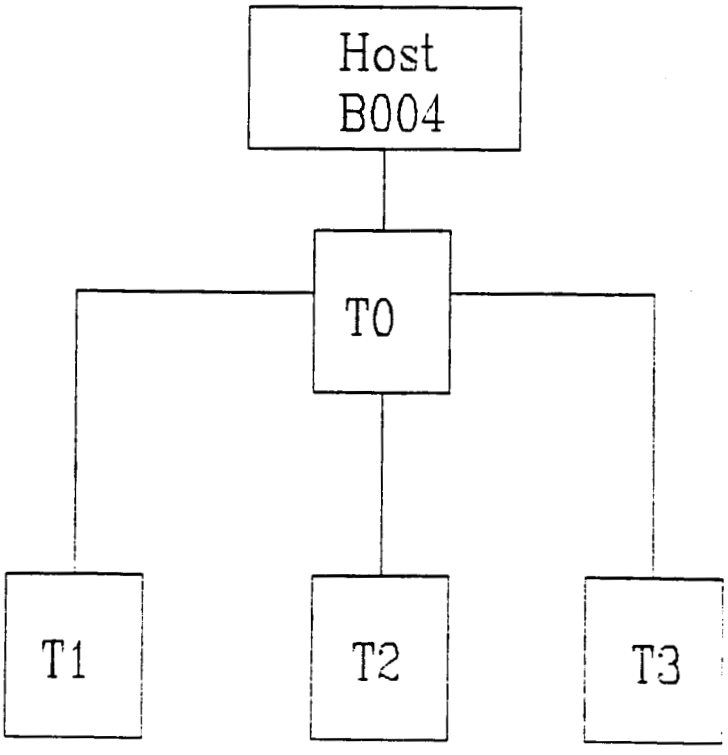


Figure 1. Fault-Tolerant Processor With a Coordinator Processor

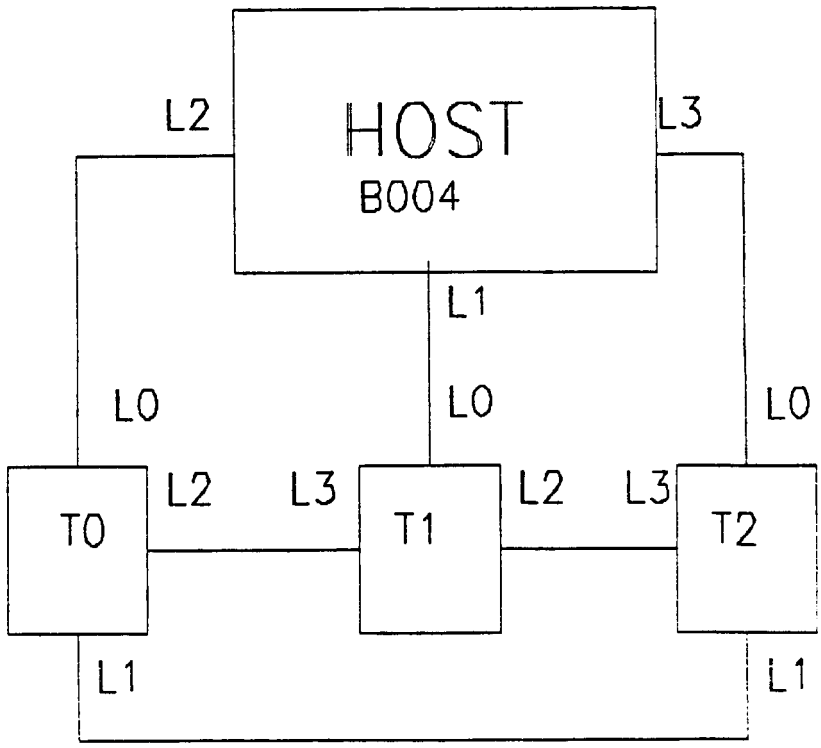


Figure 2. Transputer Fault-Tolerant Processor Architecture

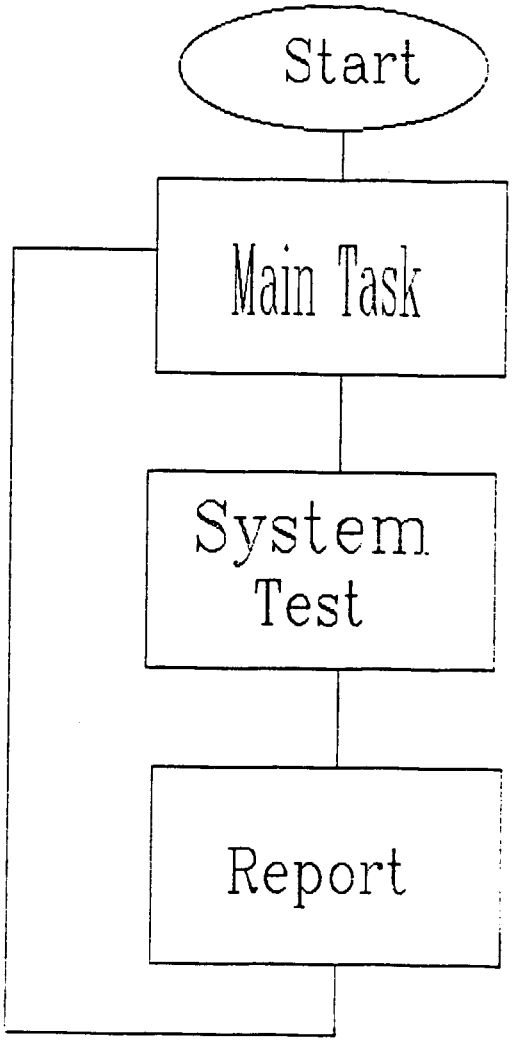


Figure 3. Program Flow Chart

The three processors are kept executing the same software sequence at the same time. As shown in Figure 3, the processors start executing the main task followed by the system testing algorithm and finally ending the sequence with the system status report that is sent to the host processor. The software to implement the sequence was written in Occam.

A graphic display was developed to facilitate the identification of any problem by the user. The system shows a constantly updated screen detailing the status of each processor and the result of the tests performed between the processors. Figure 4 illustrates the screen graphic display when all processors are on line in normal operation, as it shows processor T0 is reporting on the screen the status of T1 and T2 (the other two processors) and since everything is normal at the moment these processors are "ok".

One of the faults that the processor can detect is a software fault, where the processor on test for some reason does not get the correct answer for a numeric operation. As it can be seem on Figure 5 processors T1 and T2 found that processor T0 has a software fault and they display the occurrence of that fault on the status of processor T0. A similar software fault is simulated in processor T1 and the results are shown in Figure 6.

Another fault that can be simulated in this system, is a communication fault or hardware fault. As shown in Figure 7 a fault has occurred in a link at point a, and Figure 8 shows that due to this fault the host processor could not receive the status report from processor T2. Also, processors T0 and T2 acknowledge that a fault has been detected on the mentioned processor.

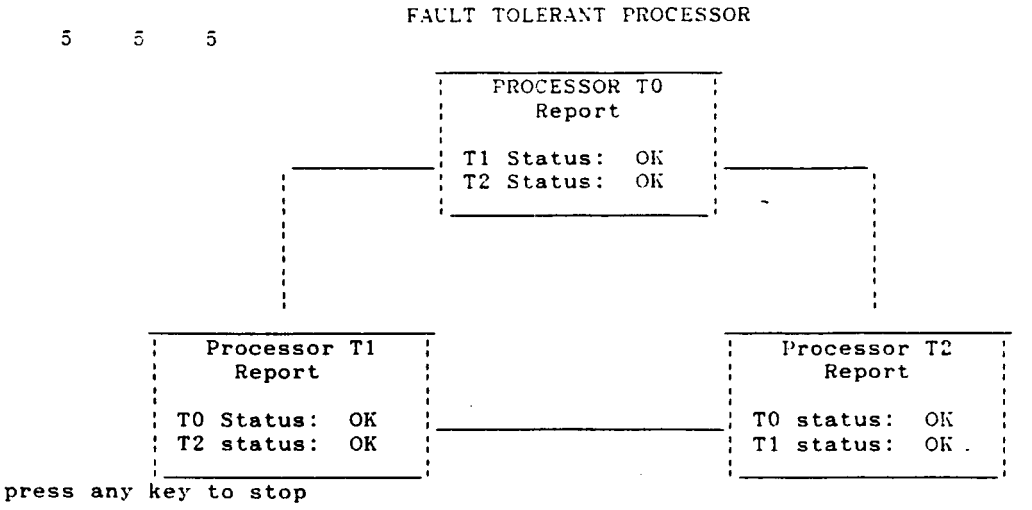


Figure 4. Transputer Fault-Tolerant Processor in Normal Operation

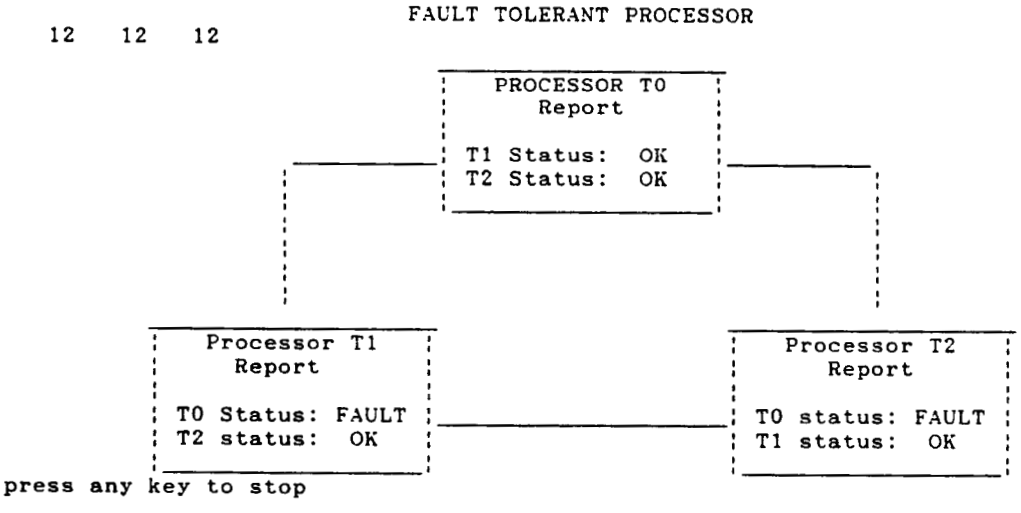


Figure 5. Software Fault in Processor T0

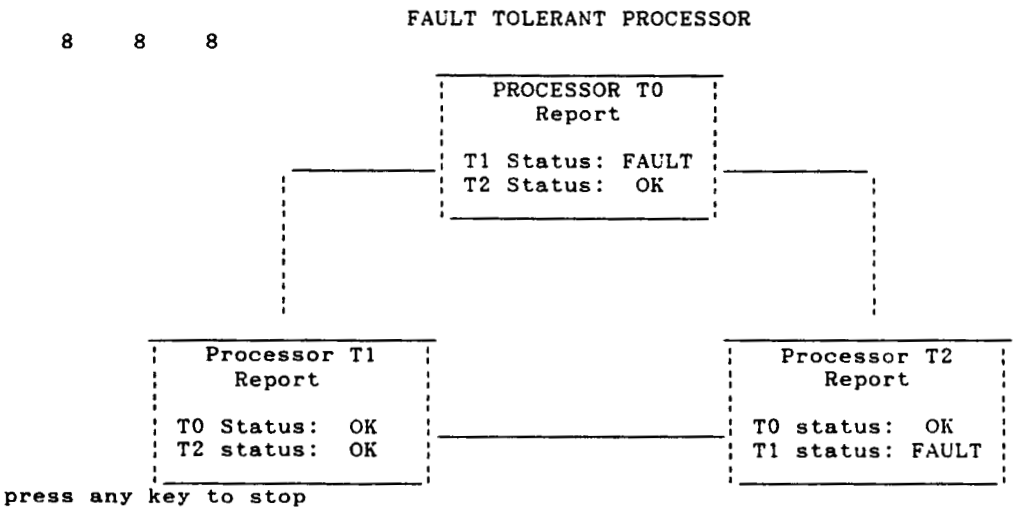


Figure 6. Software Fault in Processor T1

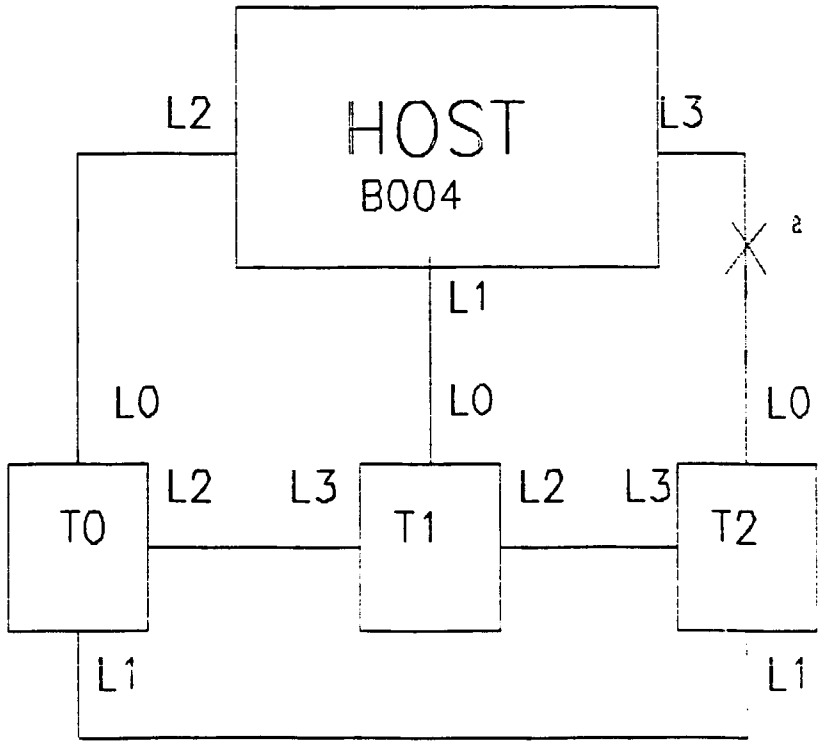


Figure 7. Fault in a Link at Point a.

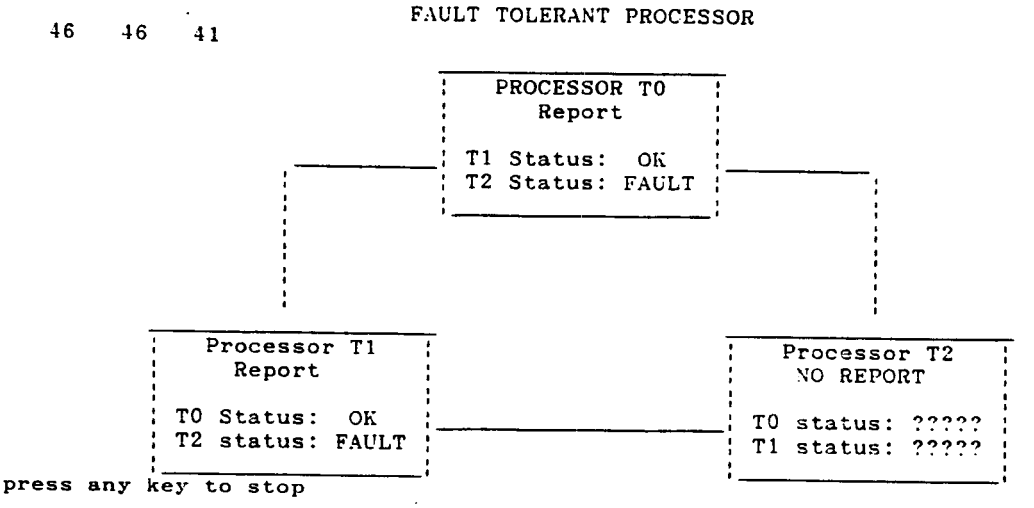


Figure 8. Hardware or Communication Fault in Processor T2.

Conclusions

The Transputer Fault-Tolerant Processor has shown to be an excellent alternative when a reliable processor is needed. More research has to be done to improve link communications, its synchronization, and link resetting after a hardware fault occurs. However the TFTP is potentially faster than other fault-tolerant processors due to the Transputer parallel processing capacity and its specially designed Occam language to facilitate concurrent processing.

References

[1]. Daniel P. Siewiorek, Robert S. Swary, "The Theory and Practice of Reliable System Design," Digital Press, c 1982.

[2]. Jacob A. Abraham, et al,"Fault Tolerance Techniques for Systolic Arrays", IEEE Trans. Computers, July 1987, pp. 65-74.

[3]. Roger Allan,"Technology Report for Fault-Tolerant Computing, Software is Finding a Powerfull Ally in Hardware",Electronic Design,October 1985,pp.111-116.

[4]. L. Beaudet, F. Eshragh, "System-level diagnostics Troubleshoot Multiprocessors", Computer Design, July 1987, pp. 77-82.

[5]. J. Bond,"Parallel-Processing Concepts Finally Come Together in Real Systems",Computer Design,June 1987, pp.51-74.

[6]. Advanced Information Processing System (AIPS), Demostration Notes, The Charles Stark Draper Laboratory,Inc,Cambridge,Massachusetts.

RANDOM LOADING FATIGUE CRACK GROWTH — CRACK CLOSURE CONSIDERATIONS

Final Report

NASA/ASEE Summer Faculty Fellowship Program – 1987

Johnson Space Center

Prepared by:	Keith Ortiz, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	University of Arizona Department of Aerospace & Mechanical Engineering Tucson, Arizona 85721
NASA/JSC	
Directorate:	Engineering
Division:	Structures and Mechanics
Branch:	Materials
JSC Colleague:	Royce G. Forman
Date:	August 14, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

The prediction of fatigue crack growth is an important element of effective fracture control for metallic structures and mechanical components, especially in the aerospace industry. The prediction techniques available and applied today are mostly based on fatigue crack growth measurements determined in constant amplitude testing. However, while many service loadings are constant amplitude, many more loadings are random amplitude. This report is concerned with an investigation to determine which statistics of random loadings are relevant to fatigue crack closure. This report briefly reviews the fundamentals of random processes and crack closure, then it qualitatively discusses the relevance of certain random process parameters to the crack closure calculation. A course for further research is outlined.

1. INTRODUCTION

The growth of fatigue cracks to fracture is the most critical life-limiting criterion for almost all metallic structures and mechanical components. This is especially true of aerospace structures and components because of the high stresses they experience. The risk posed by fatigue and fracture is significant and its management is known as fracture control. Two elements essential to effective fracture control are the control of the initial sizes of cracks, flaws, or crack-like defects, and the prediction of the growth of fatigue cracks.

Initial crack sizes are typically controlled by manufacturing procedures and inspections. For example, to make a fastener hole one might either drill the hole directly, or first drill a smaller pilot hole and then ream it. The initial flaw sizes characteristic of the two processes might be expected to be different. The fastener holes would then be subjected to nondestructive inspection. There are several techniques commonly used, each has limitations on the size flaw that can be detected with confidence. Together the initial flaw size distribution and the flaw detectability define the probability distribution of flaw sizes that go into service.

For fracture control purposes, the largest initial flaw size expected to pass through manufacturing and inspection into service is chosen as the starting point for fatigue crack growth calculations. In order to facilitate accurate and consistent calculations, NASA has developed and distributed a computer code called FLAGRO [Forman et al., 1986]. This program automatically computes the increments of crack growth for the specified loading, and continues the calculations until the crack reaches its final critical size and shape. The fatigue loading is specified by a "loading spectrum," which is a tabulation of the number of stress cycles occuring in different stress ranges.

Standard loading spectra have been developed for various applications, for example, the NASA Goddard Space Flight Center's spectrum developed for payloads carried by the Space Shuttle. The growth increment for a given stress cycle is determined by "linear damage accumulation theory," which means that any effects related to potential interactions between fatiguing loads are ignored.

Load interaction effects have been extensively documented for overloads introduced into constant amplitude crack growth tests. It is observed that a large tensile overload will cause a reduction in the crack growth rate for some time after the overload. This behavior is accounted for by crack closure theory which states that an overload leaves behind residual stresses which hold the crack closed for a portion of the subsequent stress cycles. The effective amplitude of the subsequent stress cycles and the corresponding growth rate are thereby reduced. Hence, in some loadings load interaction effects may be quite important.

Many fatigue loadings, especially in the aerospace industry, are actually random processes for which the interaction effects are probably significant. So it is

not surprising that members of industry have recently asked NASA to add load interaction effects to FLAGRO's capabilities. This capability should make the crack growth predictions more accurate. Furthermore, since the dominant effect of load interactions is to prolong fatigue life, this would have a beneficial impact on future designs via weight reductions. It would also enable the certification of existing structures for longer life extensions, which is emerging as a critical issue for NASA on both the Space Shuttle and the Space Station.

Two types of approaches to the random loading problem are found in the literature: cycle-by-cycle calculations and calibrations to characteristic stresses. In the cycle-by-cycle approach, crack growth is calculated one cycle at a time with the order of cycles preserved. The crack growth rate for each cycle is determined from constant amplitude fatigue tests, and load interactions are accounted for by reducing the effective stress range by a factor related to plastic zone size, e.g., Wheeler [reference] and Willenborg [reference].

In the characteristic stress approach, the crack growth rate under random loading is re-calibrated to a characteristic value of the stress intensity factor, typically the root-mean-square (RMS) value. The sequence of loads is lost, and a new crack growth rate curve must be generated for each new loading. The crack growth rate curve is typically reported as the growth per flight, da/dF, for example.

However, before industry adopts a "nonlinear damage accumulation theory", the impact on fracture control safety margins should be assessed so that the probability of fracture is not radically changed. The currently established scheme for fracture control is based on and calibrated to a long experience base. Certain probabilities of fracture are achieved by these practices and have been accepted implicitly by the engineering community. If one element of the scheme is perturbed, other elements must also be adjusted. For example, if one could predict longer fatigue lives with greater certainty, one might be able to tolerate larger initial flaws, but the accuracy of the inspection may have to be increased in order to maintain the same probability of fracture.

Therefore, the random loading fatigue problem is really two-fold. First, a better understanding of the mechanisms of fatigue crack growth under random loading is necessary so that more accurate crack growth predictions can be made. Second, a better understanding of the true safety margins (i.e., probabilities of fracture) achieved by current practices is necessary so that new practices may be confidently introduced. This research project is concerned only with the prediction aspect of the random loading problem. Understanding the safety margins would logically be a second project.

This report examines the interplay between random loadings and fatigue crack closure as a first step in defining a research program on random loading fatigue. Section 2 provides background on random loadings and crack closure. Section 3

examines the prediction of crack closure under known variable amplitude fatigue loadings. Section 4 then extrapolates the concepts to random loadings. Section 5 presents a course for future research.

2. BACKGROUND

This section briefly reviews the basic concepts of random loadings and fatigue crack closure. Due to the limited space available, a comprehensive literature review is not attempted.

2.1 Random Loading Statistics

Figures 1a and 1b show samples of two different processes which might represent the fatigue loading on a crack. The theory of random processes is well known. We will only briefly headline the important concepts below. For a more in-depth treatment a number of standard texts are available, e.g., Bendat and Piersol [1986].

The probability that a random process $X(t)$ will be found below a value x_0 at an arbitrary time t_0 is given by the marginal probability distribution function (PDF) of the process, denoted $F_X(x_0)$. The expected value or mean, $E[X(t)] = \mu_X(t)$, and the variance, $\sigma_X^2(t)$, are the most important statistics of the process and are generally time dependent. The square root of the variance is the standard deviation, $\sigma_X(t)$.

The normal (or Gaussian) distribution is commonly applied to random processes of engineering interest. It has two parameters, the mean, μ, and the variance, σ^2. Its equation is

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{(t-\mu)^2}{\sigma^2}\right] dt \tag{1}$$

A process with this PDF is called a Gaussian process.

The probability that, at two times t_1 and t_2, $X(t)$ will be found below x_1 and x_2, repectively, is given by the joint PDF, $F_{X(t_1),X(t_2)}(x_1, x_2)$. Higher order distributions may also be defined, but are not used very often. The joint PDF is used to calculate the autocorrelation function which is defined as the expected value of $X(t_1)$ times $X_(t_2)$, i.e., $R_X(t_1, t_2) = E[X(t_1)X(t_2)]$.

If the joint PDF remains constant over all time, the process is called stationary. Otherwise, the process is nonstationary. For stationary processes, the mean is constant, $\mu_X(t) = \mu_X$, the variance is constant, $\sigma_X^2(t) = \sigma_X^2$, and the autocorrelation function depends only on the length of the time difference $\tau = t_2 - t_1$, $R_X(\tau) = E[X(t)X(t+\tau)]$. Note that for $\tau = 0$, the autocorrelation function is equal to the mean-square value of the process, $R_X(0) = \sigma_X^2 + \mu_X^2$.

It is necessary to subtract the mean value of the process before any further analysis. The linearly transformed process then has a mean of zero. If necessary, the mean can always be added back. Then the mean-square value given by the autocorrelation function at $\tau = 0$ is equal to the variance, $R_X(0) = \sigma_X^2$, and the square root of the mean-square value (RMS) is equal to the standard deviation, σ_X.

The random processes of greatest engineering significance are Gaussian, stationary, and have zero-means. Once one has obtained the autocorrelation of such a process all the important statistical properties can be derived from it. These processes are assumed for the remaining discussion unless otherwise indicated.

The spectral density function, $S_X(\omega)$, is the Fourier transform of the autocorrelation function, and thus contains the same statistical information. While the autocorrelation function has great mathematical significance, the spectral density function is more useful to engineers because it gives us the distribution of energy in a random process as a function of frequency, ω. Figures 2a and 2b show the spectral density functions for the processes shown previously in Figures 1a and 1b, respectively. In Figure 2a it is easily seen that the energy of the first signal is concentrated in a very narrow band of frequencies. This is called a narrowband process and it manifests itself as a sine wave of slowly varying random amplitude and phase shift. In Figure 2b it is seen that the energy of the second signal is spread out over a wider band of frequencies. This is called a wideband process and it manifests itself as a highly irregular waveform, what is commonly thought of as "random".

From the properties of Fourier transforms, the area under the spectral density function is equal to the value of the autocorrelation function at $\tau = 0$, which is equal to the variance of a zero-mean process, σ_X^2. The square root of the area is thus the RMS of the process, equal to the standard deviation, σ_X.

The area under the spectrum is a special case of a moment of the spectral density function. More generally, the k-th moment of the spectrum is given by

$$m_k = \int_0^\infty |\omega|^k S_X(\omega) d\omega \tag{2}$$

The area is the zero-eth moment, $k = 0$. Other moments, particularly $k = 2$ and $k = 4$, are used to calculate the rate at which the process crosses different levels (level crossings), such as, $x = 0$ (zero-crossings), and the rate at which peaks occur in the process. The ratio of the rate of zero-crossings to the rate of peaking is known as the irregularity factor, α, and is a measure of the bandwidth of the process, $0 \leq \alpha \leq 1$. Note that $\alpha = 1$ is the ideal theoretical narrowband process. Also note that $\epsilon = \sqrt{1 - \alpha^2}$ and $k = 1/\alpha$ are also used as bandwidth measures.

The spectral moments are also used to determine the probability distribution of extrema, i.e., peaks and valleys. These distributions depend on α and on σ_X^2.

The author has derived approximate distributions for the height of the rise from a valley to the next peak and for the amplitude of rainflow counted fatigue cycles which depend on α and σ_X^2 as well as on other spectral moments [Ortiz and Chen, 1987]. These distributions are useful for calculating linear fatigue damage.

Not all random processes of interest to engineers are stationary. In general, nonstationary processes are much more difficult to handle and must be treated on a case-by-case basis. Standard approaches to these problems exist, but are beyond the scope of this section.

2.2 Fatigue Crack Closure

The correlation between the rate of fatigue crack growth, da/dN, and the cyclic range of the elastic stress intensity factor, $\Delta K = K_{max} - K_{min}$, is well established. However, to understand crack closure, it is necessary to reconsider the evolution of the plastic strain field as the crack grows longer. We summarize below the key elements of the theory, first proposed by Elber [1971]. For an extensive review, see Banerjee [1984].

On loading the crack from K_{min} to K_{max}, the material at the crack tip undergoes monotonic plastic straining and the crack tip extends. The zone of plastically deformed material has a width of approximately

$$r_p \approx \begin{cases} \frac{1}{2\pi}\left(\frac{K_{max}}{\sigma_y}\right)^2 & \text{plane stress} \\ \frac{1}{6\pi}\left(\frac{K_{max}}{\sigma_y}\right)^2 & \text{plane strain} \end{cases} \tag{3}$$

Note that, since K_{max} typically depends on $S_{max}\sqrt{\pi a}$, r_p depends linearly on a and quadratically on S_{max}. Figure 3 illustrates the extent of the monotonic plastic wake for constant amplitude fatigue loading.

When the crack is unloaded, the singularity at the crack tip continues to exist until the crack faces meet, i.e., until the crack is closed. The crack tip stress field reverses. However, due to the Bauschinger effect the effective stress intensity on unloading is only half as much as on loading. Thus, the reversed plastic zone is only one-fourth (i.e., one-half squared) the width of the monotonic plastic zone. Figure 3 also illustrates the extent of the reversed plastic zone.

One result of this plastic deformation is obviously crack extension. Another, less obvious, result is that there is net stretching of the plastic zone. This creates a compressive residual stress field in the plastic wake behind the crack tip. When the crack is unloaded, the crack faces meet before the load is fully removed, closing the crack. The stress at which this occurs is called the crack closing stress. On reloading, the crack remains closed until sufficient load is applied to open the crack, the crack opening stress, S_{op} (or K_{op}). Experimental evidence indicates that the

opening and closing stresses are approximately the same. The importance of crack closure to fatigue crack growth is that there is no plastic deformation at the crack tip until the crack opens. In effect, ΔK is reduced to $\Delta K_{eff} = K_{max} - K_{op}$.

Various researchers have attempted to predict S_{op} for constant amplitude loadings. The most extensive study to date is by Newman [1984], who fit the following equations to finite element model results:

$$S_{op}/S_{max} = \begin{cases} A_0 + A_1 R + A_2 R^2 + A_3 R^3 & \text{for } R \geq 0 \\ A_0 + A_1 R & \text{for } -1 \leq R < 0 \end{cases} \tag{4}$$

where

$$A_0 = (0.825 - 0.34\alpha + 0.05\alpha^2)[\cos(\pi S_{max}/2\sigma_o]^{1/\alpha} \tag{5}$$

$$A_1 = (0.415 - 0.071\alpha) S_{max}/\sigma_o \tag{6}$$

$$A_2 = 1 - A_0 - A_1 - A_3 \tag{7}$$

$$A_3 = 2A_0 + A_1 - 1 \tag{8}$$

The opening stress is seen to depend on S_{max}, the R-ratio, $R = S_{min}/S_{max} = K_{min}/K_{max}$, and σ_o, the material's flow stress, which Newman takes to be the average of the uniaxial yield stress and the uniaxial tensile strength, and on α, a "constraint" factor on tensile yielding. The material is assumed to yield when the applied stress is equal to $\alpha\sigma_o$. For plane stress, $\alpha = 1$, and for plane strain, $\alpha = 3$.

Crack closure has been used to explain the dependence of da/dN on the R-ratio, which is approximated by Forman's law

$$da/dN = \frac{C\Delta K^m}{(1-R)K_c - \Delta K} \tag{9}$$

According to the crack closure theory, da/dN should correlate with the effective range of the stress intensity factor, $\Delta K_{eff} = K_{max} - K_{op}$, which is somewhat less than the nominal $\Delta K = K_{max} - K_{min}$. Limited experimental results show good agreement.

Crack closure theory may also be used to explain the load interaction effects which occur when stress ranges of different amplitudes are intermingled in a load history. According to the theory, an occasional tensile overload in an otherwise constant amplitude stress history creates a residual stress field which increases the crack opening stress over a portion of the crack path ahead of the crack tip. This reduces ΔK_{eff} and the corresponding da/dN until the crack grows out of the affected region. This effect is known as "crack growth retardation". Similarly, an occasional compressive overload can reduce the opening stress and thereby increase da/dN. This is known as "crack growth acceleration".

The majority of theoretical and experimental work on crack closure has been done on constant amplitude loadings. In the following sections we discuss the prediction of crack closure under variable amplitude and random amplitude loadings.

3. CRACK OPENING STRESS FOR VARIABLE LOADING

The subject of crack closure under variable amplitude loadings is still very new. In the occasional overload problem, the crack opening stress obviously varies as the crack grows. In a complex variable amplitude loading, such as shown in Figure 4, the crack opening stress could vary significantly during the load history, but there are reasons why it may not. In this section we will try to understand why S_{op} might or might not vary significantly during a variable amplitude loading, and how it might be predicted.

3.1 Direct Approach

Consider a variable amplitude loading such as shown in Figure 4. Suppose this same loading block is repeated over and over. And for the moment, suppose that the crack does not grow during a single application of the block. It is evident that the maximum extent of the monotonic plastic zone is determined by the highest stress peak, S_{maxh}. The net plastic stretching is also governed by the lowest stress valley, S_{minl}. Therefore, the crack opening stress should be a function of the R-ratio:

$$R_{lh} = S_{minl}/S_{maxh} \tag{10}$$

Now let us relax the restriction that the crack does not grow during a single application of the loading block. If the crack grows "slowly" during the block, the net plastic stretching and, hence, the crack opening stress will still be controlled by R_{lh}. By "slowly" we mean the crack tip should not grow out of the plastic zone during the block. The plastic wake behind the crack tip for the variable amplitude loading should appear to be smooth, and should be the same wake as for constant amplitude loading from S_{minl} to S_{maxh}. Figure 5 illustrates the concept and shows a counter-example for which the crack does grow out of the maximum plastic zone, as for example in the occasional overload problem.

Thus, if the crack is growing slowly, S_{op} will remain constant, the plastic wake will be the same as for constant amplitude loading from S_{minl} to S_{maxh}, and predictable using R_{lh} in the relationships found in constant amplitude tests. The crack growth rate can then be predicted from ΔK_{eff}. The method described has been proposed by Schijve [1980]. The correlation of his predictions with experimental results appears promising, but is not perfect. The errors in Schijve's predictions are most probably due to using Elber's original equations for S_{op}. Perhaps Newman's more recent equations would give better results.

3.2 Iterative Approach

One potential problem with the above approach has to do with the following. In a variation of occasional overload testing, the overload is followed immediately by a second identical overload. The observed retardation is longer than for a single overload [Banerjee, 1984]. Since the plastic zone theoretically is not enlarged by a significant amount, this suggests that the second overload somehow strengthens the compressive residual stresses. One infers that the crack opening stress observed in constant amplitude tests is not necessarily the same as observed in occasional overload testing. So perhaps one should not use the highest peak and the lowest valley to define the equivalent constant amplitude R-ratio, but rather the second highest peak, or an average?

Newman [unpublished] suggests the following iterative procedure:

First iteration:

1. Find the highest peak, S_{maxh}, and the lowest valley, S_{minl}.
2. Calculate an initial S^*_{op} using S_{minl}, S_{maxh} and

$$R^* = S_{minl}/S_{maxh} \tag{11}$$

3. For each stress cycle, calculate the initial effective stress ranges

$$\Delta S^*_i = S_{max,i} - S^*_{op} \tag{12}$$

Second iteration:

4. Calculate the weighted average stress range given by

$$\overline{\Delta S} = \left[\frac{1}{N_T} \sum_{i=1}^{N_T} \Delta S^{*m}_i \right]^{1/m} \tag{13}$$

where m is Paris exponent and only "positive" cycles, totaling N_T, are counted.

5. Calculate the average effective maximum stress

$$\overline{S_{max}} = S^*_{op} + \overline{\Delta S} \tag{14}$$

6. Calculate the final effective opening stress $\overline{S_{op}}$ using the lowest valley S_{minl}, the average effective maximum $\overline{S_{max}}$ and the final effective R-ratio

$$\overline{R} = S_{minl}/\overline{S_{max}} \tag{15}$$

7. Calculate the final effective stress ranges with which to calculate ΔK_{eff}

$$\Delta S_{eff,i} = S_{max,i} - \overline{S_{op}} \tag{16}$$

Note that the first three steps correspond to Schijve's method. The next three steps represent a second iteration to get an average effective R-ratio in step 6. Step 7 is then the calculation of the effective ΔK.

3.3 Discussion

At this point it is impossible to say which of the two methods presented is better. Again, the rate at which the crack grows through the plastic zone of the "overload" is of great importance. If the crack is growing very slowly, it might take several or many blocks to get through the plastic zone. There could be a reinforcing of the plastic zone by repeated applications of the loading block. In this case the plastic wake would be similar to the constant amplitude loading wake. Schijve's approach would appear to be reasonable under these circumstances.

On the other hand, if the crack is growing so quickly or if the loading block is so long that the crack tip advances significantly during a block, there might not be any reinforcing of the plastic zone. In this case the plastic wake would not be similar to the constant amplitude loading wake. In fact, the wake would be irregular and S_{op} would vary non-monotonically. In this case, Newman's approach might be an acceptable approximation. The conditions under which either method would be valid must be established by experiment.

So far we have assumed that the same variable amplitude block is repeated over and over again. However, in real loadings each variable amplitude block would be different in many respects. In fact, the loading blocks would likely be random. This leads us to extend the discussion to random amplitude loadings in the next section.

4. CRACK OPENING STRESS FOR RANDOM LOADING

From our discussion so far it is evident that the crack opening stress, S_{op}, for a random amplitude loading will probably vary randomly. That is, S_{op} is also a random process. For a stationary random loading, S_{op} should also be stationary. For a nonstationary random loading, S_{op} could still be stationary because it should vary more slowly than the loading.

For the purpose of discussion, we first assume that the stress process, $X(t)$, is stationary. Assuming the process is Gaussian, which is often the case, the mean value, μ_X, and the variance, σ_X^2, define the marginal probability distribution. These would be standard statistics of the process which should always be available. The

spectral density function and the bandwidth, e.g., α, should also be readily available. The general question is, "How do these parameters affect crack growth and S_{op}?" In particular, we ask how ΔK and R-ratio might be measured for stationary random loadings. We later discuss the treatment of nonstationary random loadings, some of the practical considerations in designing a testing program, and aspects of the calculations of the statistics of random loading fatigue crack growth.

4.1 Random Loading Stress Ranges

As with constant amplitude fatigue crack growth, we should expect influences of ΔK and R-ratio. Taking ΔK first, a natural question is, "Which should da/dN correlate with, ΔK_{mean} or ΔK_{rms}?" Note that these two parameters are actually the expected values of ΔK raised to the powers one and two, respectively. That is,

$$\Delta K_{mean} = E[\Delta K] \tag{17}$$

and

$$\Delta K_{rms} = E[\Delta K^2]^{1/2} \tag{18}$$

Because it depends on a random process, the crack growth rate under random loading is also random. To correlate experimental data, we should probably be looking at the average or expected growth rate, $E[da/dN]$. Assuming the Paris fatigue crack growth law is valid, at least in first approximation, we have $da/dN = C\Delta K^m$. Thus, the expected value of da/dN is

$$E[da/dN] = CE[\Delta K^m] \tag{19}$$

We should thus expect da/dN to correlate with

$$\Delta K_m = E[\Delta K^m]^{1/m} \tag{20}$$

where m is dependent on the material.

Hibberd and Dover [1977] show that this gives better correlations than ΔK_{mean} or ΔK_{rms}. Barsom [1973] reports that the rms value works well for steels, for which it happens that $m \approx 2$; this also fits the hypothesis.

The distribution of the cyclic ranges, ΔK, must be known in order to calculate the expected value of ΔK^m. The approximate distribution of rises/falls or the approximate distribution of rainflow cycles derived by the author [Ortiz and Chen, 1987] could be used for this distribution. These distributions are known to depend on the spectral shape, especially on the irregularity factor, α.

There is still some question as to which cycle counting method is applicable to fatigue crack growth. For example, Schijve [1980] uses the rising ranges, but Sunder

et al. [1984] present fractographic evidence favoring rainflow. This is an important point that needs to be resolved.

4.2 Random Loading R-Ratio

For a random loading R-ratio, $Q = K_{mean}/K_{rms}$ and $\gamma = S_{mean}/S_{rms}$ have been proposed and are essentially the same. One might think of Q or γ as measuring the number of standard deviations the mean is above zero. A large Q or γ (e.g., greater than 3) implies that the probability of $X(t)$ crossing below zero is small. Hence, these ratios are indirect measures of the ratio of lowest valley to highest peak, R_{lh}.

We could propose more direct measures of the random R_{lh}. For instance, we could define

$$R_\beta = \frac{\mu_X - \beta\sigma_X}{\mu_X + \beta\sigma_X} \tag{21}$$

Assuming a Gaussian process, $R_{\beta=1.28}$ would be, for example, the ratio of the load which $X(t)$ is above 90% of the time and below 90% of the time. However, for the same loading this would take on different values depending on the percentage level indicated by β.

A more meaningful and stable ratio is the ratio of the average valley to the average peak, $R_{\overline{min},\overline{max}}$

$$R_{\overline{min},\overline{max}} = \frac{E[S_{min}]}{E[S_{max}]} \tag{22}$$

If the process is Gaussian, this calculation would use the well known distributions for peaks and valleys.

Keep in mind however, that the thing we really wish to measure is the maximum plastic zone width, which we think governs the crack opening stress. Since the maximum width is caused by extreme values of the peaks and valleys, we should be looking for a statistic defined by the extremes, rather than the averages. The distinction could be significant if the distributions of peaks and valleys are highly skewed or asymmetical. (A single parameter indicating the skewness is the ratio of the "largest" peak to the average peak.) So perhaps the ratio should be of the 90% valley to the 90% peak, for example. But, we would encounter the same difficulty regarding the percentage level as we have with R_β above. Furthermore, at this point in time, we do not know for what percentage level we should set the ratio.

A third approach might be taken. From the definition of the constant amplitude R-ratio, $R = S_{min}/S_{max}$, one can show

$$R = 1 - \frac{\Delta S}{S_{max}} \tag{23}$$

For random loadings we must substitute for ΔS and S_{max}. Again, either the average values, $\overline{\Delta S}$ and $S_{\overline{max}}$, or the $\beta\sigma$ values, e.g., 90% values, could be used, with the same difficulties associated with β as before. And the question of which method to count cycles, ΔS, needs to be resolved.

Other definitions of R could easily be proposed. A main goal of any experimental investigation should be establish which R-ratio is valid.

4.3 Nonstationary Loadings

We now extend our scope to nonstationary random loadings. Many of the loadings of interest to us are nonstationary, e.g., there are ground-to-air-to-ground (GAG) cycles for aircraft and changes in vibration level for payloads during launch. This creates two new problems: the probability distributions indicated above are no longer valid, and the crack opening stress may no longer be stationary. The treatment of nonstationary processes must be on a case by case basis. However, we can address these problems in general terms.

Regarding the probability distributions, we are still interested in the distributions of cycles, peaks and valleys. However, the underlying distribution of $X(t)$ should not be expected to be Gaussian. It is likely that the highest peaks and lowest valleys are associated with "deterministic", i.e., non-random, phenomena. For example, the lowest valley in a GAG cycle may correspond to the plane at rest on the ground, while the highest load may correspond to a certain flight maneuver. It is evident that one must study the physics of each situation in order to understand these distributions. These distributions would then by applied as before if the opening stress is still stationary.

Regarding the stationarity of the crack opening stress, it is obvious that in the occasional overloading problem the crack opening stress process is nonstationary. That is, the distribution of the process changes with time. The cause of this is the sudden change in the plastic wake width associated with the overloadings, such as illustrated in Figure 5. If overloads are extremely rare, as is the case in occasional overload testing, the analysis appears to become very complicated. However, if the overloads happen on a fairly regular basis, such as, once every flight, then the crack opening stress process might still be considered stationary, if the crack grows slowly between overloads. This is probably the case for most variable amplitude loadings of practical interest. (Besides, if the rarely occuring overload is a possibility, one probably would not want to depend on it occuring and causing retardation for a fatigue life prediction, unless the occurence of the overload could be controled and made to happen.)

Thus, the analysis for most nonstationary loadings would be very similar to that for stationary loadings, except for a little more effort. The question that needs

to be resolved is how slow the crack has to grow in order to maintain a smooth plastic wake.

4.4 Practical Considerations for Testing

There are a number of practical considerations which must also be addressed if one is to design a testing program for random loadings. These include the truncation level and the programmed load sequence.

Theoretically, the Gaussian process $X(t)$ has no upper or lower bounds, i.e., it may go to infinity in either direction. Naturally, this does not happen in practice and cannot be reproduced in the laboratory. In laboratory tests, the peak loadings are truncated to some finite level which is often specified by the ratio of the maximum stress to the RMS, called the clipping ratio or crest factor. It has been shown that truncating the highest loads leads to faster crack growth, due to lower opening stresses. Because of the critical impact of the largest peak on crack closure, the truncation level should always be explicitly stated. Unfortunately, many researchers do not recognize how important this is and do not state it, which makes their results ambiguous. Even though the choice of truncation level has been shown to be very important, we have no rational way of choosing it at this time. This definitely is something to be studied.

Another practical consideration has to do with programmed load sequences. In many laboratory tests, the load ranges are arranged non-randomly in a programmed sequence either from smallest to largest, or vice versa. This could have a significant impact on crack closure. The sequence of loads should be followed as faithfully as possible in a test, rather than rearranging them in such programmed sequences.

4.5 Probability Calculations

Finally, regarding the calculation of the statistics of fatigue crack growth, i.e., the mean and variance of the time to failure, there are several things to note.

First, one should keep in mind that it is the residual stress behind the crack tip that controls the opening stress. So S_{op} acting for a certain load cycle is determined by loads acting some time previously. Since it is likely to take a reasonably long time for the crack to grow through the plastic zone, it is likely that the stress process acting on the crack tip would be statistically independent of the concurrent crack opening stress process. Independence of the two processes leads to certain mathematical simplifications.

A possible exception to this is for highly narrowband loadings which maintain high correlations between peak loads for long times. It is possible to imagine situations for which the concurrent stress and opening stress processes are also highly correlated (either positively or negatively). The assumption of independence should be made with care.

Second, this is a highly nonlinear problem. Therefore, the average opening stress would not necessarily lead to the average fatigue life prediction. If we want to directly calculate the average fatigue life using a constant value for S_{op}, we would need to use a value other than the expected value. This implies that we might expect large (potentially unconservative) biases from deterministic calculations of fatigue life.

Third, there is a stochastic modeling question near and dear to the author's heart. Should the opening stress process be modeled as a stochastic process evolving in time, $S_{op}(t)$, or in space along the crack path, $S_{op}(a)$? Since the opening stress is to be mathematically manipulated with the applied stress process to get the effective ΔS, we would prefer the time model, $S_{op}(t)$, for computational simplicity. On the other hand, since the opening stress is actually related to the residual stresses behind the crack tip, the space model, $S_{op}(a)$, might be preferred on physical grounds. This is the sort of thing professors and graduate students think about ....

5. CONCLUSIONS

Regarding the prediction of random loading fatigue crack growth using crack closure for fracture control analysis, we make the following observation.

- Current fracture control procedures which apply linear damage accumulation theory neglect crack closure. In this approach, the theoretically applied load spectrum is normalized so that the largest peak is equal to the maximum design load, i.e., the maximum design stress is seen once each time the spectrum is applied. Since the maximum design load should occur only rarely, perhaps never, this is thought to be "conservative", more linear damage is calculated than would be expected. On the other hand, if crack closure theory is applied, the maximum design load reduces the rate of growth by increasing the crack opening stress. Repeated application of this load would be unconservative! Veers et al. [1987] demonstrate the unconservativeness of repeated applications of such a programmed loading.

Regarding the course of future research, we make the following suggestions for theoretical and experimental work.

- There is a basic need for understanding the statistical nature of common random loadings. Are they stationary? Are they nonstationary? What are the distributions of cycles, peaks and valleys? From these statistics we would like to be able to predict whether or not crack closure is likely to have a significant impact on the fatigue life.
- The correlation of fatigue crack growth rate data with random loading equivalents for ΔK and R, e.g., ΔK_m and $R_{\overline{min},\overline{max}}$, should be experimentally established for stationary Gaussian processes. Different spectral bandwidths

and spectral shapes should be investigated. (Beware that extra care should be taken with narrowband spectra because of the persistance of high correlations between peaks.) The loadings should be chosen so that the maximum theoretical differences in crack growth rate are calculated using the different candidates for ΔK and R. This includes differences due to cycle counting methods.

- The random loading ΔK and R, concepts should then be extended to nonstationary random loadings. Experimental correlations should be made with the particular loading spectra of interest.
- Truncation levels should be investigated. What is the relationship between the highest peak and the second highest peak? Can general rules regarding the truncation level and the amount of retardation lost be established?
- An interesting idea which deserves exploration is the visualization of the plastic wake. It would be facinating and informative to correlate crack growth rate with the structure of the wake, for instance as depicted in Figure 5. How might the wake be seen? Photographs against grids have been used to show gross plastic deformation. Some sort of surface treatment or etching might show the actual extent of plastically deformed material in finer detail.
- Finally, the impact of adopting crack closure models into fatigue crack growth analysis for fracture control should be invesitgated. Using a more accurate crack growth analysis for random loadings places a greater burden on inspection to screen initial flaws. The confidence in the inspection should be greater. How much greater? An effort should be made to quantify the probability of fracture using current fracture control practices and using crack closure. The theoretical probability of fracture should be kept constant, unless a change can be justified.

6. REFERENCES

Banerjee, S., "A Review of Crack Closure," Materials Laboratory, Air Force Wright Aeronautical Laboratories, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, AFWAL-TR-84-4031, April 1984.

Barsom, J.M., ASTM STP 536, 1973, pp. 147–167.

Bendat, Julius S., and Piersol, Allan G., *Random Data, Analysis and Measurement Procedures*, John Wiley & Sons, 1986.

Elber, W., in *Damage Tolerance in Aircraft Structures*, ASTM STP 486, American Society for Testing and Materials, 1971, pp. 230–242.

Forman, R.G., et al., "Fatigue Crack Growth Computer Program NASA/FLAGRO," Materials Branch, Structures and Mechanics Division, NASA Lyndon B. Johnson Space Center, Houston, Texas, JSC–22267, August 1986.

Hibberd, R.D., and Dover, W.D., "The Analysis of Random Load Fatigue Crack Propagation," *Fracture 1977, Fourth International Conference on Fracture*, Waterloo, Canada, ICF4, Vol. 2, June 1977, pp. 1187–1194.

Newman, J.C., Jr., "A Crack Opening Stress Equation for Fatigue Crack Growth," *International Journal of Fracture*, Vol. 24, 1984, pp. R131–R135.

Newman, J.C., Jr., personal communication with R.G. Forman on May 20, 1987, unpublished.

Ortiz, Keith, and Chen, Nobel K., "Fatigue Damage Prediction for Stationary Wideband Random Stresses," *ICASP5, Fifth International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Vancouver, Canada, May 1987.

Schijve, Jaap, "Prediction Methods for Fatigue Crack Growth in Aircraft Material," *Fracture Mechanics: Twelfth Conference*, ASTM STP 700, American Society for Testing and Materials, 1980, pp. 3-34.

Sunder, R., Seetharam, S.A., and Bhaskaram, T.A., "Cycle Counting for Fatigue Crack Growth Analysis," *International Journal of Fatigue*, Vol. 6 No. 3, 1984, pp. 147–156.

Veers, Paul S., Winterstein, Steven R., Nelson, Drew V., and Cornell, C. Allin, "Variable Amplitude Load Models for Fatigue Damage and Crack Growth," submitted for review to, *ASTM Symposium on Fatigue Loading Spectra*, 1987.

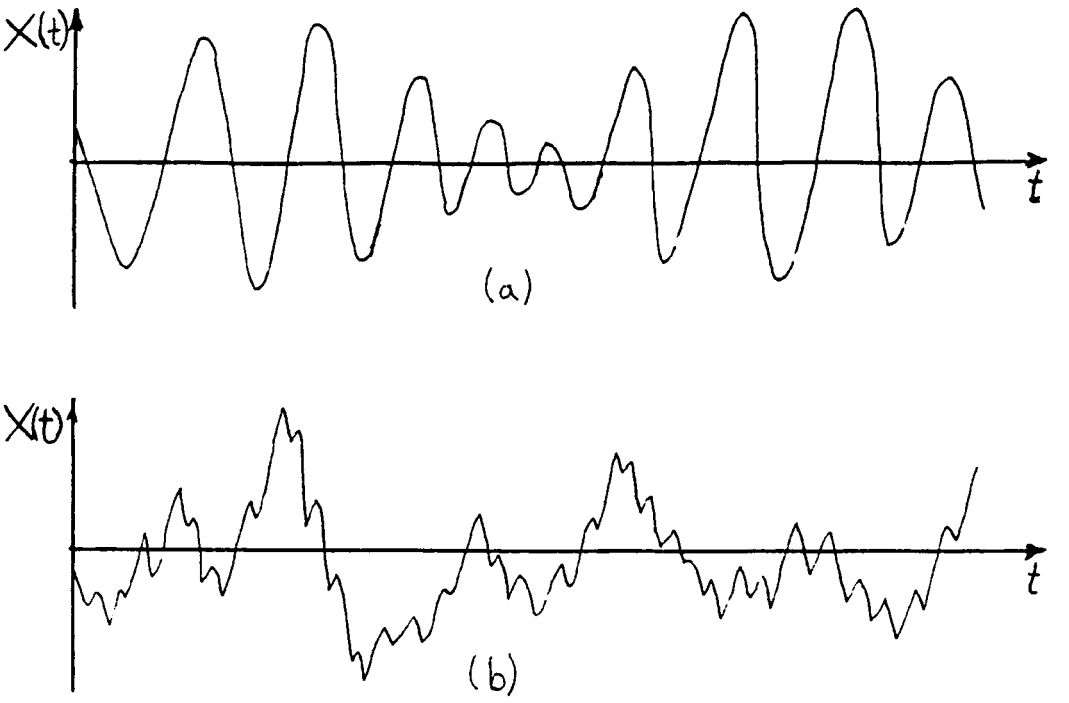


Figure 1. Sample random processes: (a) narrowband; (b) wideband.

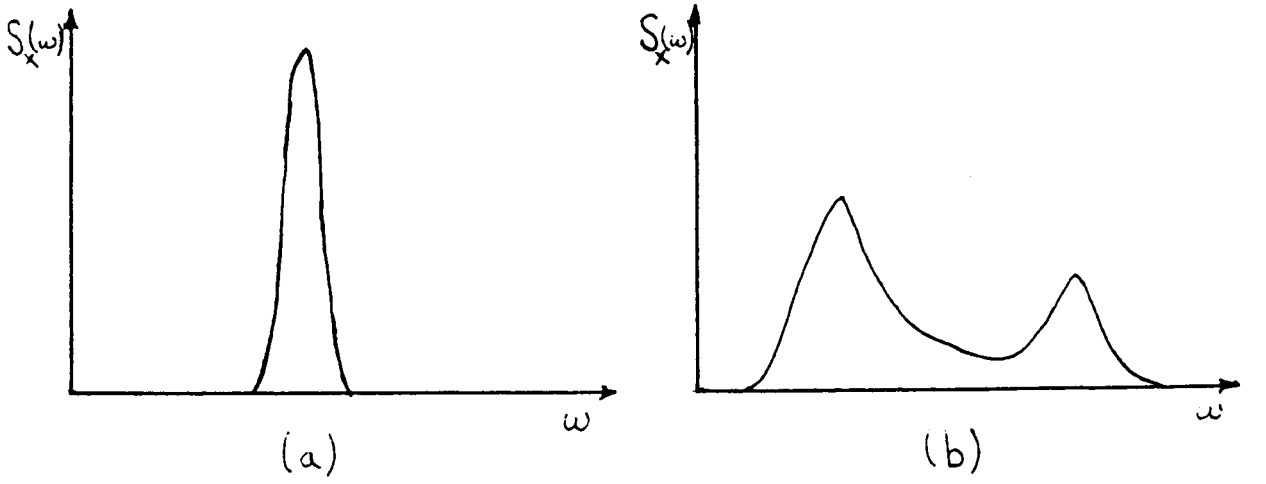


Figure 2. Spectral density functions (one-sided) for processes in Figure 1: (a) narrowband; (b) wideband.

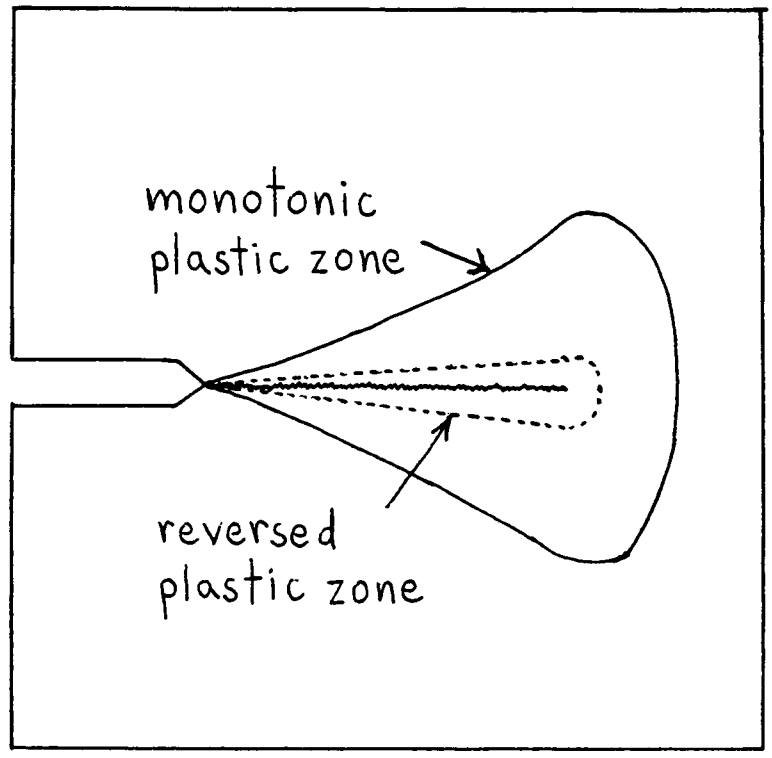


Figure 3. Illustration of the plastic wake for constant amplitude loading showing the envelope of monotonic plastic zone and the reversed plastic zone.

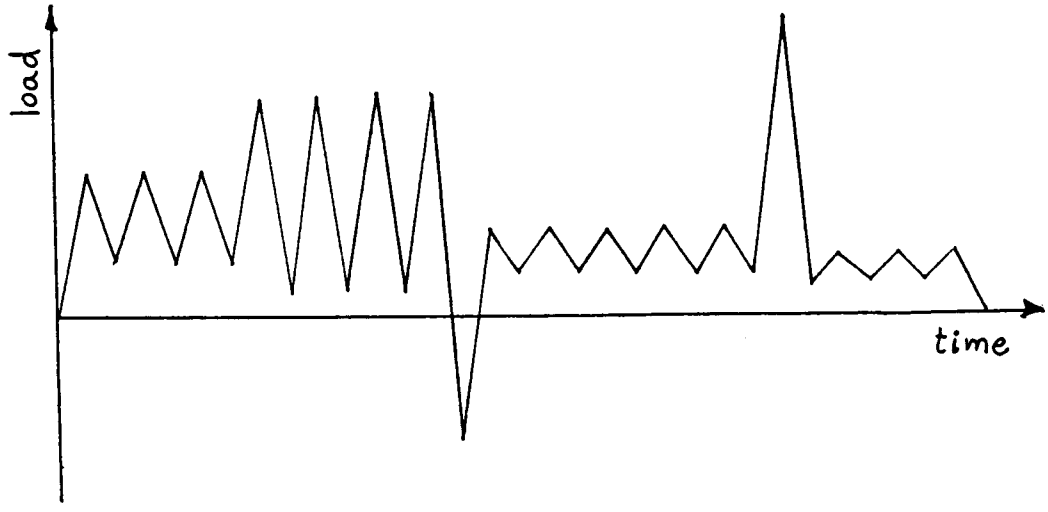


Figure 4. Example variable amplitude fatigue loading.

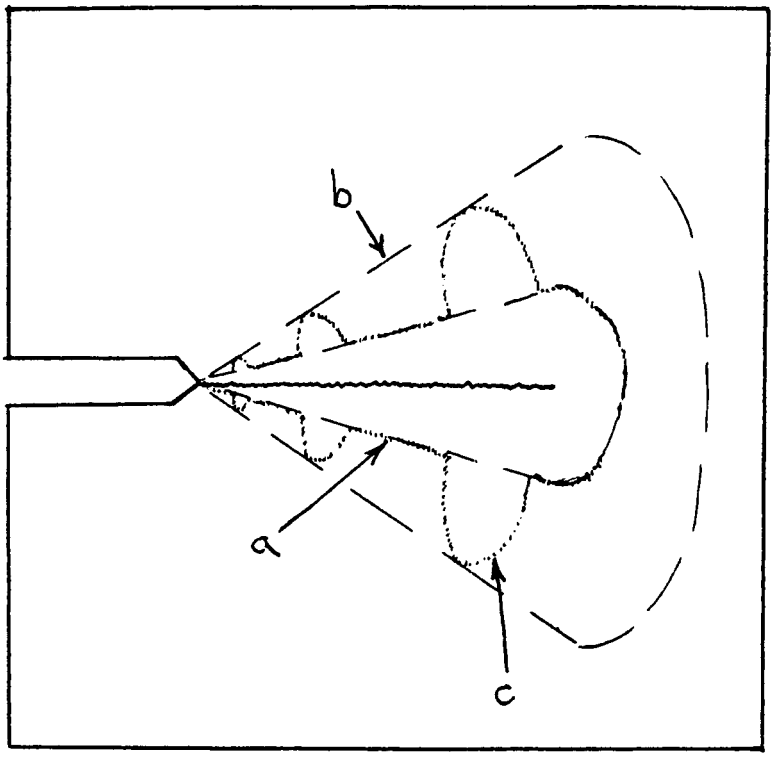


Figure 5. Illustration of the plastic wake for: (a) constant amplitude loading from 0 to S_1; (b) constant amplitude loading from 0 to S_2 greater than S_1; (c) constant amplitude loading from 0 to S_1 with occasional overloads to S_2.

TOOLS AND TECHNOLOGIES FOR EXPERT SYSTEMS A HUMAN FACTORS PERSPECTIVE

Final Report

NASA/ASEE Summer Faculty Program-1987
Johnson Space Center

Prepared by:	Navaratna S. Rajaram
Academic Rank:	Associate Professor
University Department:	University of Houston- University Park Department of Industrial Engineering Houston, TEXAS 77004

NASA/JSC

Directorate:	Space and Life Sciences Directorate
Division:	Man-Systems Division
Branch:	Crew Station Branch
Section:	Crew Interface Analysis Section
JSC Colleague:	Marianne Rudisill, Ph.D.
Date:	August 14, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

It is widely recognized that technologies based on artificial intelligence, (AI) especially expert systems can make significant contributions to the productivity and effectiveness of operations of information and knowledge intensive organizations such as NASA. At the same time, these being relatively new technologies, there is the problem of transfering technology to key personnel of such organizations. This report addresses itself to the problems of examining the potential of expert systems and of technology transfer in the context of human factors applications.

One of the topics of interest was the investigation of the potential use of expert system building tools, particularly NEXPERT as a technology transfer medium. Two basic conclusions were reached in this regard. First, NEXPERT is an excellent tool for rapid prototyping of experimental expert systems, but not ideal as a delivery vehicle. Therefore, it is not a substitute for general purpose system implementation languages such as LISP or C. This assertion probably holds for nearly all such tools on the market today. Second, an effective technology transfer mechanism is to formulate and implement expert systems for problems which members of the organization in question can relate to. For this purpose, the **LI**ghting **E**n**G**ineering **E**xpert (LIEGE) was implemented using NEXPERT as the tool for technology transfer and to illustrate the value of expert systems to the activities of the Man-Systems Division.

0. INTRODUCTION AND STATEMENT OF PROBLEMS

Introduction

It is unlikely that there ever will be a universally satisfactory and at the same time comprehensive definition of artificial intelligence (AI). However, for purposes of the present document AI will be defined as the discipline devoted to the study of human reasoning capabilities. In particular, AI is concerned with formalisms for human problem solving and decision making. It can be said that to some extent all computer programs do some decision making and problem solving. What distinguishes AI programs is their ability to reason with *symbolic data.* For instance, AI programs can manipulate algebraic expressions in symbolic form, process textual information in English, reason with linguistic and other grammatical rules, use strategic information in games of skill such as chess and these are among systems that fall under the scope of AI.

Among the branches of AI expert systems have captured the greatest attention, among professionals as well as with the general public. If AI is defined as the study of problem solving and decision making, an expert system can be defined as a computer program which can solve problems and make decisions in a specific domain. In particular, there exist expert systems which can diagnose diseases, design electronic circuits, locate faults in telephone cable networks, and perform other similarly skill demanding tasks. It can thus be seen that despite their relatively recent origins, expert systems have already addressed a fairly wide range of applications. In spite of their diversity, expert systems exhibit certain characteristic features which can be summarized as follows.

1. Expert systems are *knowledge based;* i.e., an expert system achieves its often impressive problem solving power from specialized knowledge, rather than from any general form of reasoning skill.

2. An expert system has no pre-programmed algorithm built in. On the other hand, it has to select and assemble a sequence of "elementary units of knowledge" which usually depends upon the particular problem case at hand.

These and other related issues will be elaborated upon in later sections.

Statement of Problems

The research reported herein was intended to address the following specific issues, all of them of interest to the personnel of the Man-Systems Division of NASA Johnson Space Center.

1. **Expert systems for man-systems applications.** To invstigate the relevance and usefulness of the technology of expert systems within the broad context of human factors and human-machine systems, *with the help of concrete examples drawn therefrom.*

2. **Approaches and tools.** To evaluate the scope, advantages and limitations of different approaches and tools used for expert systems. In particular, to study the use of the expert system tool NEXPERT.

3. **Technology transfer.** To use problems of interest in the man-systems area as the medium for technology transfer, with the eventual goal of enabling expert systems to be integrated into such activities.

4. **Future needs and objectives.** To assess the future needs and objectives in man-systems applications as regards tools, environments, and applications in the context of relevant problems.

Each of these issues will be described in the remainder of this report.

Conclusions

Two basic conclusions were reached which can be expected to have some bearing on future activities of the Man-Systems Division. Because of the importance which NASA attaches to AI and expert systems, they are discussed in some detail in the last section of this report. But those conclusions can briefly be summarized as follows.

1. AI and expert systems technologies can make a significant contribution to human-systems tasks and activities.

2. NEXPERT is an excellent tool for building experimental systems.

1. APPROACHES AND TOOLS

1.1 Structure of Expert Systems

As previously noted, expert systems are *knowledge based;* i.e., they solve problems by making inferences using *specialized knowledge* in a particular domain. This knowledge is usually expressed in the form of IF... THEN... rules. Such rules are called *production rules* in the technical literature. Thus, every expert system must have the following two kinds of knowledge.

- o *Domain knowledge* usually expressed in the form of production rules.

- o *Method knowledge or inference knowledge* about how to use domain knowledge.

It is convenient to think of domain knowledge as "know-what" and method or inference knowledge as "know-how."

In order for an expert system to function as a complete integrated system, it needs the following components.

1. Domain knowledge.

2. Inference mechanism.

3. A suitable computer representation of the problem state; i.e., a *data structure* representation of the problem. The usual term for this data strucure is *working memory.*

4. User interface which has the responsibility of converting the user stated problem statement into a computer readable data structure, i.e., the initial working memory.

Figure 1 shows the structure and organization of an expert system schematically.

The working of an expert system is best described in terms of a *cycle.* A cycle in an expert system consists of the following sequence of steps.

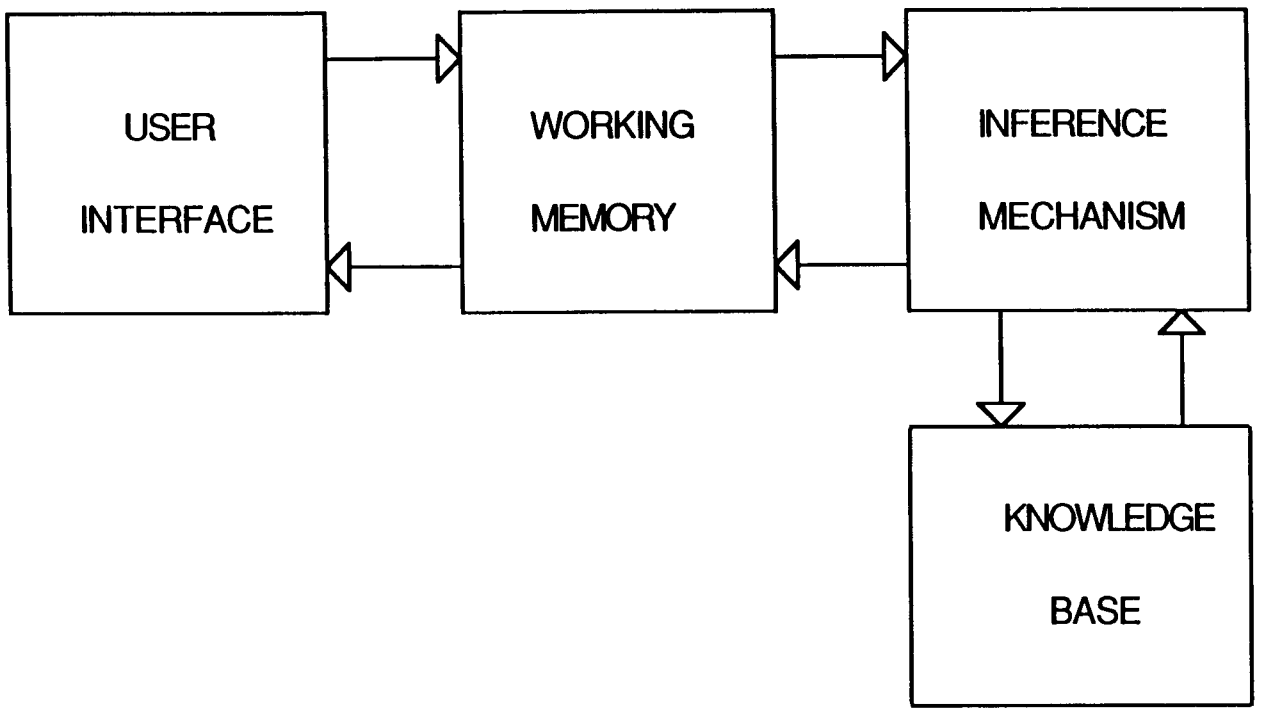


Figure 1: Organization of an expert system

1. User inputs the problem statement via the user interface. The user interface converts this input into a data structure which becomes the initial working memory.

2. The inference mechanism selects a rule from the knowledge base and applies it to the working memory. This rule application changes the working memory.

3. The inference mechanism checks to see if this changed working memory corresponds to a solution state. If it does, the system reports the solution back to the user via the user interface. If it does not correspond to a solution state, the inference mechanism selects another rule and the cycle repeats.

This is an oversimplified description of the problem solving process in most expert systems. Invariably there are factors such as search efficiency and other related matters which have to be taken into consideration in even moderate sized systems. Nevertheless, it can be said that the problem solving mechanism in an expert system essentially consists of searching through the knowledge base to find the right sequence of rules in the context of the working memory.

There are basically two ways of carrying out inference in an expert system-- *hypothesis driven* and *data driven.* The hypothesis driven inferencing is also known as *backward chaining*, while the data driven approach is called *forward chaining.* In the hypothesis driven approach, the system postulates a solution and then checks to see if the working memory justifies the hypothesis. In the data driven approach, the reverse approach is taken, i.e., a hypothesis to match the working memory is searched for. Recalling that production rules in the knowledge base have the following structure, the terms forward and backward chaining are obvious.

IF <data> THEN <conclusion (or hypothesis)>

1.2 Expert System Tools: NEXPERT

Design and implementation of an expert system to work in a specific application domain calls for the integration of the following kinds of knowledge.

1. Knowledge of the problem domain or domain expertise.

2. Knowledge of the expert system design methodology, at least in the context of the problem(s) of interest.

3. Knowledge of the implementation medium, i.e., the hardware/software environment.

There are basically two choices in the the selection of implementation software.

- o A general purpose AI oriented language such as LISP or Prolog.

- o A special purpose expert system building tool such as NEXPERT.

But no matter which approach is taken, the software and the implementor(s) must be able deal with all three of the above mentioned parts of building an expert system. In addition, the user interface part of the system might become the critical component when a system has to be delivered to end-users. This report is primarily concerned with the use of the high level expert system tool NEXPERT. The relevance of general purpose languages such as LISP is discussed in the next section. Issues relating to user interfaces do not form a central concern of this research and will merely be touched upon in the final section.

Description of NEXPERT

NEXPERT is a fairly typical expert system building tool, its distinguishing feature being a quite flexible inference engine. Unlike several other tools on the market, NEXPERT permits the user to use either forward or backward chaining, or even a combination of the two. As is the case with nearly all such tools, NEXPERT is a rapid prototyping tool for building experimental expert systems. Unfortunately, like most pioneering software products, its documentation and instructions leave something to be desired, being often vague and at times downright misleading. Only some of the technical features of NEXPERT as they relate to the problem of interest are discussed here. A general introduction to the use of NEXPERT for implementing expert systems is given in a separate document. (See "An Introduction to Using NEXPERT.")

As previously noted, any tool must provide facilities for implementing the following parts which go to make an expert system: the workingmemory, the knowledge base, and the inference mechanism. With NEXPERT, the structure of the working memory is *implicit;* i.e., it is created when the rules making up the knowledge base are defined. In addition, the user has the freedom to choose any kind of inference, even with the same knowledge base. Thus, the primary focus of NEXPERT, at least from the user's point of view has to be on the use of its syntax for expressing production rules. Details of NEXPERT syntax are discussed elsewhere. (op. cit.) It suffices to note here that the following steps are involved in building the knowldge base in NEXPERT.

1. Analysis of the problem.

2. Formulation of the problem in the format of production systems.

3. Expressing the problem solving knowledge as production rules.

4. Coding these rules in NEXPERT.

A point to be noted is that neither NEXPERT nor any other software has the design knowledge needed to build an expert system. That knowledge has to be part of the system designer's expertise. This point needs some emphasis since not infrequently one hears such claims being made on behalf of expert system building tools, including NEXPERT. Such however is not the case.

1.3 Use of General Purpose Languages

As previously noted, any general purpose language can be used as a medium for building expert systems. The choice of the language is nearly always determined by the needs of the environment in which the finished system is required to run. If the expert system under consideration is required to interface with existing databases and other external files, it is best to choose the implementation language with that consideration in view. As is invariably the case with information systems (of which expert systems are a special case) it is the structure of the data that has to interface to the problem which dictates the design including the choice of the language.

This basic fact, namely that data are by far the most important part of any information system should decide the choice of the language for expert systems also. In the latter case, this requirement translates into defining the strcture and the needs of the working memory which is nothing more than a computer representation of the problems of interest.

Once this design and implementation decision is made, the steps involved in building an expert system are the following.

1. Choice of the data structure needed to represent the problem states.

2. Formulation of the knowledge in the format of production rules.

3. Choice of the appropriate inference mechanism(s).

4. Coding of the production rules and the inference mechanism.

As can be seen, there is some additional work involved in using a basic language such as LISP instead of a special purpose expert system building tool such as NEXPERT. But the additonal work involved in coding the inference mechanism should not be exaggerated. On the other hand, there is greater flexibility in the choice of data structures for the working memory, thereby ensuring that the tool does not limit the choice of applications. In particular, existing tools can almost never be used for delivering finished expert systems to end-users. These are among the factors which have to be taken into consideration in choosing a particular language or tool for implementation. Summarizing, it can be said that NEXPERT, like others in its genre is suitable for building experimental expert systems, but not as a delivery vehicle.

2. CASE STUDY: AN APPLICATION IN HUMAN FACTORS

The objective of the case study chosen was twofold. First, to explore and demonstrate the use of the expert system technology, and in particular, the usefulness of the tool NEXPERT for building experimental expert systems. Second, to use this particular case study as a means for technology transfer to the Man-Systems Division in the sense to be described later.

2.1 Description of Expert System LIEGE

LIEGE (**LI**ghting **En**G**i**neering **E**xpert) was the expert system designed and implemented using NEXPERT to demonstrate the feasibility of both the technology and the tool (NEXPERT) for practical problems of relevance to human factors. In particular LIEGE recommends the appropriate light intensity in international units (Lux) for a specific task in a specific area of interest. Thus, LIEGE has the following capability.

Input: TASK AREA
TASK DESCRIPTION

Output: RECOMMENDED INTENSITY

Limitations of space preclude the inclusion of a detailed description in the present document. Such a description can be found elsewhere. (Ref. 1) Here, it suffices to know that the knowledge base in NEXPERT is organized according to the taxanomy given in table 1.

LIEGE uses a forward chaining or data driven inferencing mechanism. The "data" input to LIEGE are AREA and TASK which are chosen from the options menu which forms part of its knowledge base. Thus the user does not have to know which are the permissible inputs and which are not. The user however does need to know how to load the corresponding knowledge base which has been called LIEGE1.KB.

Intensity is only one of the parameters which go into the lighting design task. LIEGE has been designed with possible future extensions in mind. When fully implemented, its knowledge base will be according to the structure in figure 2.

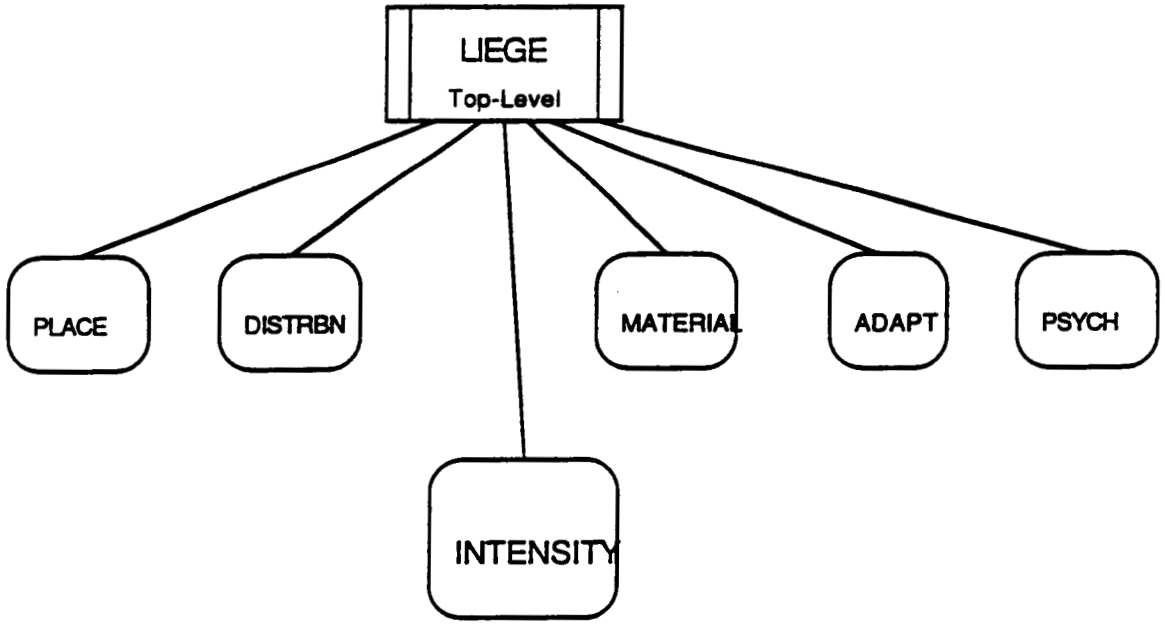


Figure 2: Potential modules in lighting design

Table 1

AREA	TASK
General	All (i.e., ambient lighting)
Translation paths	Hatches Handles Ladders
Wardroom & galley	Reading Recreation Dining Food preparation
Personal hygiene	Grooming Shower Health care
Waste management	All
Crew quarters	Reading Emergency lighting
Health maintenance	First aid Surgical I.V. treatment Exercises
Emergency lighting	Unspecified

As previously noted, LIEGE is a production system. Thus the taxanomy given in table 1 has to be converted into production rules and then encoded in NEXPERT. Here is an example rule.

IF *AREA is health maintenance*
and TASK is surgical
THEN *INTENSITY 920 lux is confirmed*

2.2 Application Development for Technology Transfer

At this time, expert systems is a technology in transition; i.e., it is a technology whose potential for contributing to the efficiency and productivity of knowledge intensive organizations is recognized, but is yet to realize that potential. A basic problem in implementing expert systems for specific applications is what is known as the *knowledge acquisition bottleneck*.

Knowledge Acquisition Bottleneck

From the discussion so far, it is evident that designing and implementing an expert system for a specific application calls for the analysis of the problem in fine detail, and then expressing the skills needed to solve the problem, i.e., *domain knowledge* in the format of AI, in particular as production rules. At this time, the great majority of domain experts lack design knowledge about expert systems. As a result, there exists a communication gap between technologists who know how to build expert systems, and application specialists with domain knowledge, i.e., between those with "what knowledge" and those with "how knowledge." This is known as the *knowledge acquisition bottleneck.* This has to be addressed whenever we want to exploit a new or emerging technology such as AI. Once the existence of this bottleneck is recognized, one can take either of the following two approaches in attempting to overcome it.

- Familiarize knowledge engineer with the domain knowledge to the point that he/she acquires sufficient depth in the latter turn this knowledge into expert systems; i.e., turn the knowledge engineer into a domain expert.

- Train the domain expert in the use of tools and techniques of building expert systems; i.e., turn him/her into a knowledge engineer capable of generating expert systems from his/her own domain knowledge.

Experience has shown that the second approach is by far the more effective and realistic. In this approach, an AI specialist functions as an instructor and a consultant on specific problems for which expert systems are being built by domain experts. This was the approach followed during

the present project as the means of technology transfer. Obviously, such technology transfer has to take place before any organization can efffectively use a new technology like AI and expert systems. Thus, the following course appears to be the most productive and at the same time the most pragmatic means of technology transfer.

- o Identify problems of interest within the organization, problems which the personnel in the organization can relate to.

- o Train and assist domain specialists in these areas to turn part of their expertise into expert systems. This calls for the assistance of at least one experienced knowledge engineer, but the "leverage" from this expertise will be high.

The second of these approaches was followed during the project reported here. This was greatly assisted by the lighting design problem which workers in the division could relate to. Suggestions for furthering this initiative are made in the next section.

3. CONCLUSIONS AND SUGGESTIONS FOR THE FUTURE

The goals set at the beginning of the present project were met or exceeded. Nevertheless, the project is best regarded as the beginning of a new direction rather thas as the culmination of an effort. With this perspective, the following conclusions and recommendations are presented for future reference.

Conclusions

1. NEXPERT is an excellent tool for building experimental expert systems, especially useful for rapid prototyping. The learning period needed to put NEXPERT to practical use is quite short. Further, its versatile inference facilities make it especially attractive.

When all these factors are taken into consideration, NEXPERT is at least as good as any other tool on the market, even though many of them cost several times as much as NEXPERT and require expensive and hard to use special purpose workstations.

2. At the same time, NEXPERT should be seen as a tool for building experimental systems and prototypes rather than as a delivery vehicle. In this regard, it is not a substitute for LISP or C.

By this observation, it is not at all being suggested that NEXPERT is "worse" than LISP, Prolog or C for implementing expert systems, but that it is a completely different vehicle, being a special purpose tool rather than a general purpose language. In other words, it is a case of the proverbial comparison between "apples and oranges."

3. There are many problems which fall under the category human-machine systems which can potentially benefit from the infusion and integration of AI and expert systems technologies. One of the more intriguing and potentially useful applications is the area of expert system monitoring and design of crew safety.

Suggestions for the Future

The conclusions just presented suggest possible options for the future. In particular, they indicate what needs to be done so that AI and expert systems can move beyond the experimental stage and begin to make material contributions to the relevant operations and tasks at hand.

1. Additional problems in the human-machine systems area should be identified and formulated in the framework of AI and expert systems.

2. Tools such as GCLISP and GoldWorks which enable expert systems to be built and delivered for end users should be used for the purpose so that actual systems can be delivered and their benefits evaluated.

The steps just outlined can make a significant contribution towards incorporating this new technology into existing operations. In this regard, general purpose tools such as LISP and C can be used to build and deliver systems to end users. These tools should be chosen with the specific goals in view; i.e., the application should drive the tool selection and not the other way, as is too often the case. In additon, the following technology alternatives are worth examining for possible future use.

1. Advanced computer architectures for learning machines, such as the ACA and the Connection Machine architectures.

2. Object-oriented-programming for general purpose prototyping and for building integrated user environments.

3. Expert systems with "linguistic" rule selection formalisms for dealing with uncertain situations.

The final technology option above is a crucial advance in which basic research is currently being done and has considerable future potential for the design of more flexible systems. In particular, systems with such linguistic capabilities instead of simple pattern match in the inferencing stage can lead to systems a step beyond the current state of the art. Such expert systems can deal not only with uncertain situations, but also with unforeseen circumstances, possibly even exhibiting some "common sense."

The goal of common sense reasoning by machines has so far proved somewhat elusive. It is the belief of the present author that the linguistic approach just outlined offers an alternative to logic based approaches currently being investigated. Although this is clearly a basic research area, it seems likely that such research might yield practical benefits surprisingly soon. This and other technologies are worth exploring for potential benefits.

Acknowledgements: The author is grateful for the opportunity for working with the Man-Systems Division, particularly to Dr. Marianne Rudisill. He is also grateful to Mr. Charles Wheelwright for his generous help with the knowledge base for LIEGE. Finally, he is grateful to Ms. Jeri Brown for generous help with the resources and support.

4. REFERENCES

1. Rajaram, N.S. (1987) *An Introduction to Using NEXPERT,* NASA Johnson Space Center (In preparation).

2. Rudisill, M.L. and Burns, M.J. (1986) *Human Factors of Expert Systems,* NASA Johnson Space Center.

3. *NEXPERT Fundamentals,* Neuron Data Inc., 1985.

4. *Man-Systems Integration Standards vol.1,* NASA-STD-3000, Johnson Space Center.

5. Charniak, E.J. and McDermott, J. (1984) *Introduction to Artificial Intelligence,* Addison-Wesley.

AUTOMATIC ANTENNA SWITCHING DESIGN FOR EXTRA VEHICULAR ACTIVITY (EVA) COMMUNICATION SYSTEM

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Manjit S. Randhawa, Ph. D.
Academic Rank:	Professor
University and Department:	Southern University College of Engineering Baton Rouge, Louisiana 70813
NASA/JSC	
Directorate:	Engineering
Division:	Tracking and Communications
Branch:	Electromagnetic Systems
JSC Colleagues:	George Dickey Arndt, Ph. D. John C. D. Ngo
Date:	August 14, 1987
Contract:	NGT 44-001-800

ABSTRACT

An Extra Vehicular Activity (EVA) crewmember has two-way communications with the Space Station in the Ku-band frequency (12-18 GHz). The maximum range of the EVA communications link with the Space Station is approximately one kilometer for nominal values for transmitter power, antenna gains, and receiver noise figure. The EVA Communications System, that will continue to function regardless of the astronaut's position and orientation, requires an antenna system that has a full spherical coverage. Three or more antennas that can be flush mounted on the astronaut's space suit (EMU) and/or his propulsive backpack (MMU), will be needed to provide the desired coverage.

As the astronaut moves in the Space Station, the signal received by a given EVA antenna changes. An automatic antenna switching system is needed that will switch the communication system to the antenna with the largest signal strength. A design for automatic antenna switching is presented and discussed.

INTRODUCTION

The EVA crewmember will be working on the Space Station, examining and repairing different systems and components. He will be communicating with the base station using his transceiver in the Ku-band frequency. A reliable communication link between EVA and Space Station needs to be established for various positions and orientations of the EVA crewmember. A spherical coverage will be desired to establish the reliable communication link. A multiple wide-beamwidth antenna system is needed to provide the desired spherical coverage. The number of antennas needed and their optimum positions where they can be flush mounted on the crewmember's life support system (EMU) and/or his propulsive backpack, need to be studied. Mr. John C. Ngo of the Electromagnetic Systems Branch at NASA/JSC has studied various antennas for the Ku-band and their radiation patterns. He plans to study the number of antennas needed and their optimum locations on the mock-up model of the space suit and the backpack in the very near future.

The strength of the signal received by the EVA antenna from the Space Station will change as the crewmember moves. When the signal strength received by a particular EVA antenna falls below a certain threshold level, there is a need to switch to the EVA antenna that has the strongest received signal. An automatic antenna switching system is desired to accomplish this task. A scheme for this

switching is shown in Figure 1 . The top part of Figure 1 shows a transmitter(TX)/receiver(RX) module for the EVA radio connected through a circulator to a microwave single-pole multi-throw PIN diode switch, which has been connected to four different omni antennas. When the signal strength received by a certain EVA antenna falls below a threshold as indicated by the AGC level, the signal processor which is a programmed computer, will send a command signal to a microwave sampling switch that has been connected to 10 dB directional couplers that are coupled to the omni antennas. The microwave PIN diode sampling switch that is shown in the bottom part of Figure 1 will start sampling the received signals from various antennas. These samples in the - 110 to - 70 dBm power range will pass through a 13.4 - 13.7 GHz bandpass filter (BPF), a low noise amplifier (LNA), and a Ku-band mixer that will down convert the samples into the 450 - 750 Mhz UHF range. This 300 MHz wide range has seven different 40 MHz wide channels assigned to different users. A suitable BPF (40 MHz) will pick out the right channel for the user and pass it to a second mixer that will convert it down to 70 MHz IF signal in the 50 - 90 MHz range at - 80 to - 30 dBm power level. This 40 MHz wide channel consists of three subchannels, a 22 MHz wide live video channel in the return link, a 300 KHz wide low rate video or voice or data channel, and a 800 KHz wide data rate channel in the forward link. The synthesized local oscillator will pick out the low data rate channel. The IF signal is passed through a 70 MHz bandpass filter, that was designed and built, and is then fed to a signal level detector. The block diagram of this signal

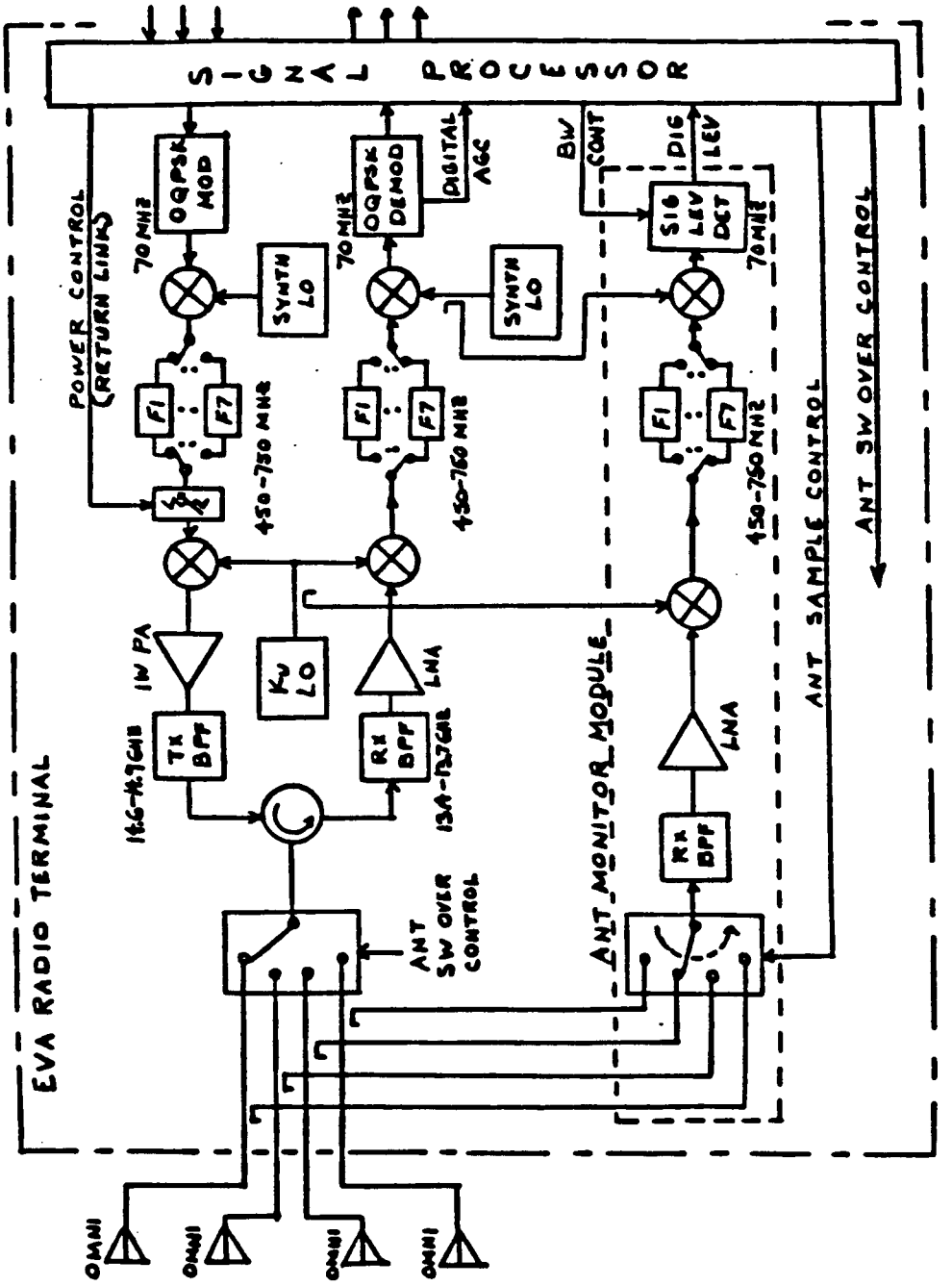


Figure 1. EVA KU RADIO TERMINAL ANTENNA SWITCHING CONCEPT

level detector is shown in Figure 2. It consists of a 54 dB cascaded amplifier, a peak detector, a 25 dB dc amplifier, and an A/D converter. The circuit for the peak detector - dc Amp. - A/D converter is shown in Figure 3. The output of the A/D converter is fed to the signal processor that would be programmed to compare the strength of the signal samples from various antennas. The signal processor will then issue a command signal for the single-pole multi-throw switch and cause it to switch to the antenna with the largest signal strength.

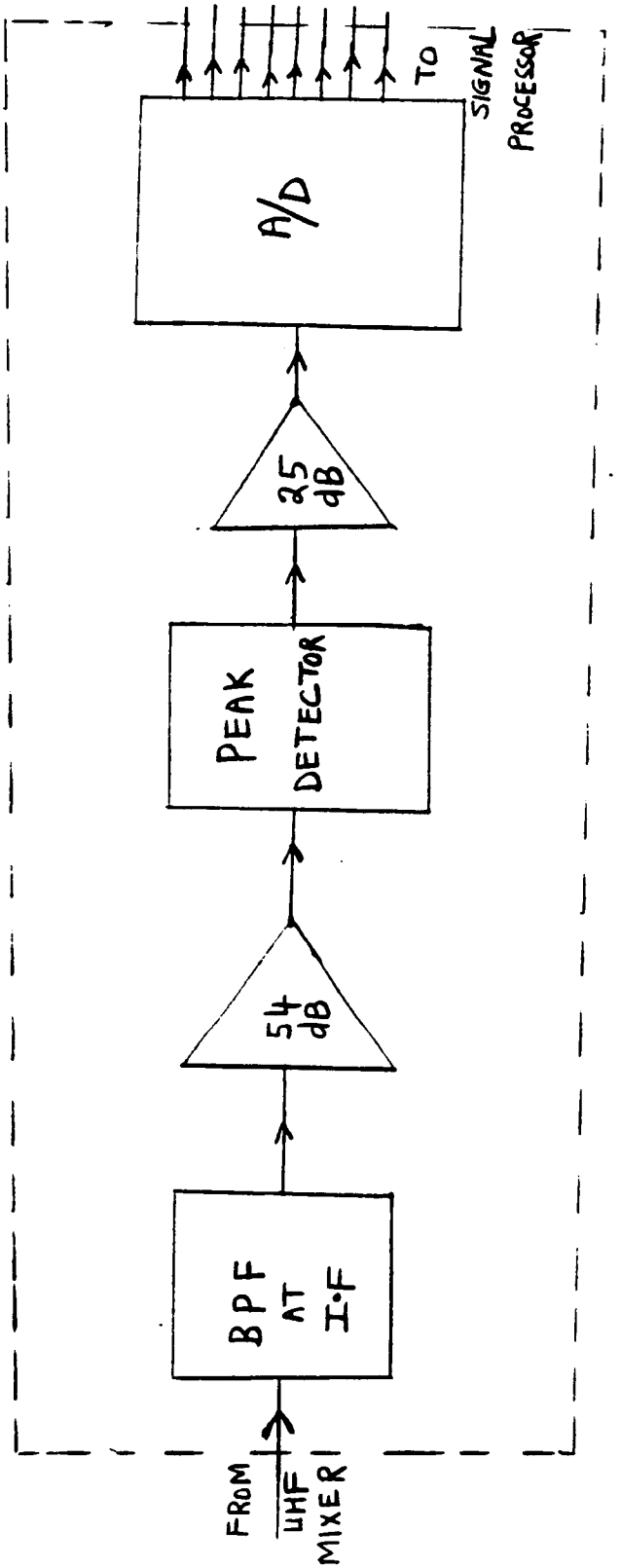


Figure 2. Signal Level Detector Block Diagram

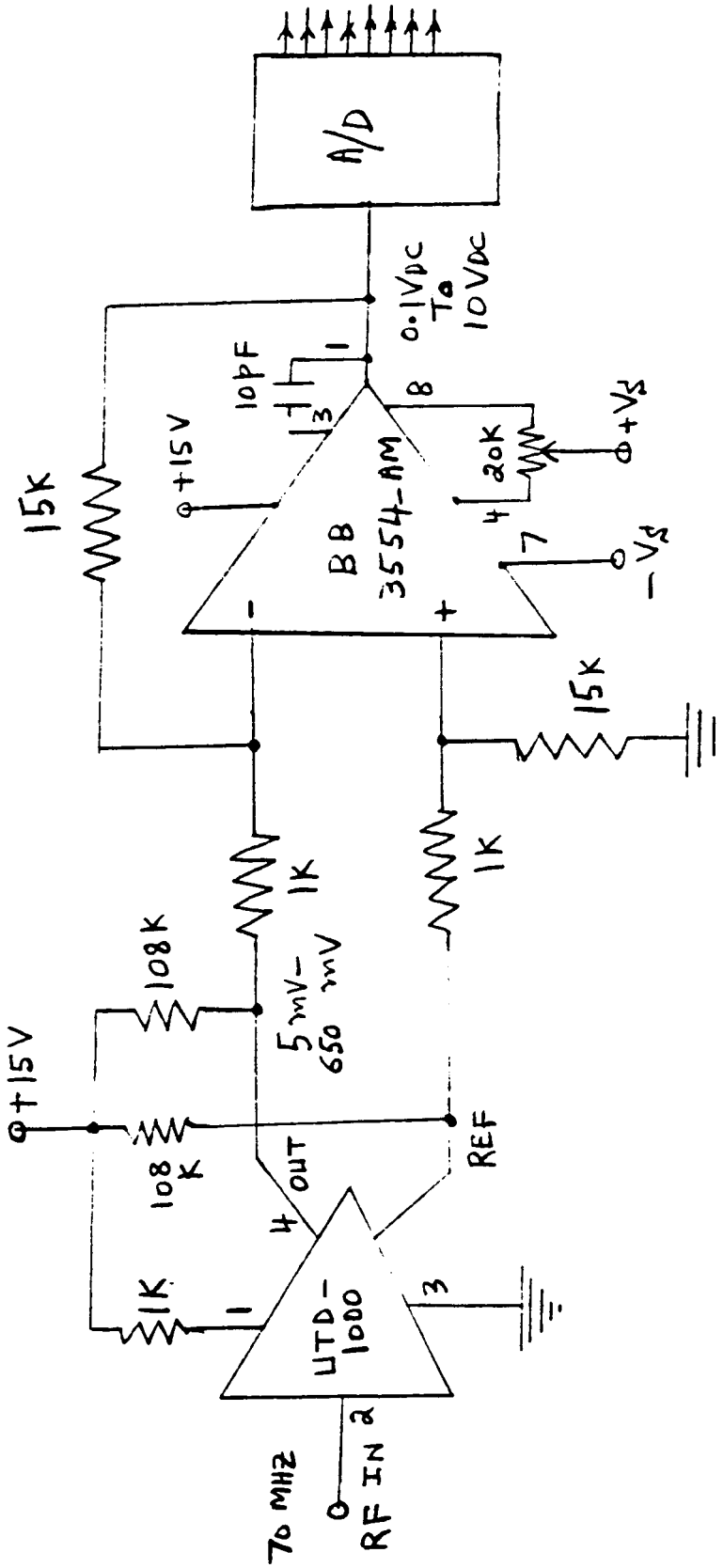


Figure 3. Peak Detector - Inverting Amp - A/D Converter Circuit

FINDING

The implementation of the design of the automatic antenna switching system required selecting suitable PIN diode switches, designing and testing the signal level detector, selecting the directional couplers and ordering the other needed hardware components. The antenna signal monitoring module of Figure 1 between the sampling switch and signal level detector uses the same type of blocks as the receive module just before the demodulator. The transmit/receive module board for EVA was built by RCA and since an extra board was available, it would be used for monitorong antenna signals. The block diagram of the signal level detector is shown in Figure 2. Its input is the signal from UHF mixer. A brief description of various components in the signal level detector is given below.

Band Pass Filter: To pass the low or high data rate subchannel at 70 MHz with 300 or 600 KHz bandwidth required a band-pass filter. This filter circuit was designed, bread-boarded and tested. The design goals were:

Center Frequency = 70 MHz

Bandwidth = 1.5 MHz

Pass Band Ripple = 0.1 dB

Filter Attenuation < - 15 dB at f>75 Mhz

A Chebyshev design with three LC sections was chosen as it

had steeper roll off and it met the design goals. The filter circuit was synthesized using ESYN program in the EESOF package. The synthesized circuit is shown in Figure 4. A file for a microwave CAD program called TOUCHSTONE was created for the circuit synthesized. The TOUCHSTONE circuit file is shown in Figure 5. The TOUCHSTONE program simulates the circuit and plots the frequency response which is shown in Figure 6 and Figure 7. The circuit response meets the design goals. The filter circuit was optimized by TOUCHSTONE using Monte Carlo analysis to meet the performance goals of:

Pass Band Loss > - 5 dB

Out of Band Loss < - 15 dB at f> 75 MHz

The Monte Carlo analysis indicated that to meet the performance goals, the component tolerance has to be very tight, less than 2 percent. A printed circuit board layout was made using MICAD program. A rubylith mask (1:1) was made using MICMASK program and a printed circuit board was etched in the Lockheed Lab. The components were mounted on the board using split-tip welding and soldering techniques. The circuit was tested using the HP83508 Sweep Oscillator and the HP8757A Scalar Network Analyzer and the desired response was not obtained. Different values were tried for the series inductors and the variable series Giga-trim Capacitors (.5-2.5PF) were tuned so that the series section resonates at 70 MHz. The series inductors that gave close to satisfactory results were 1.5 and 1.6 microhenrys. In the shunt branch, the chip capacitor with

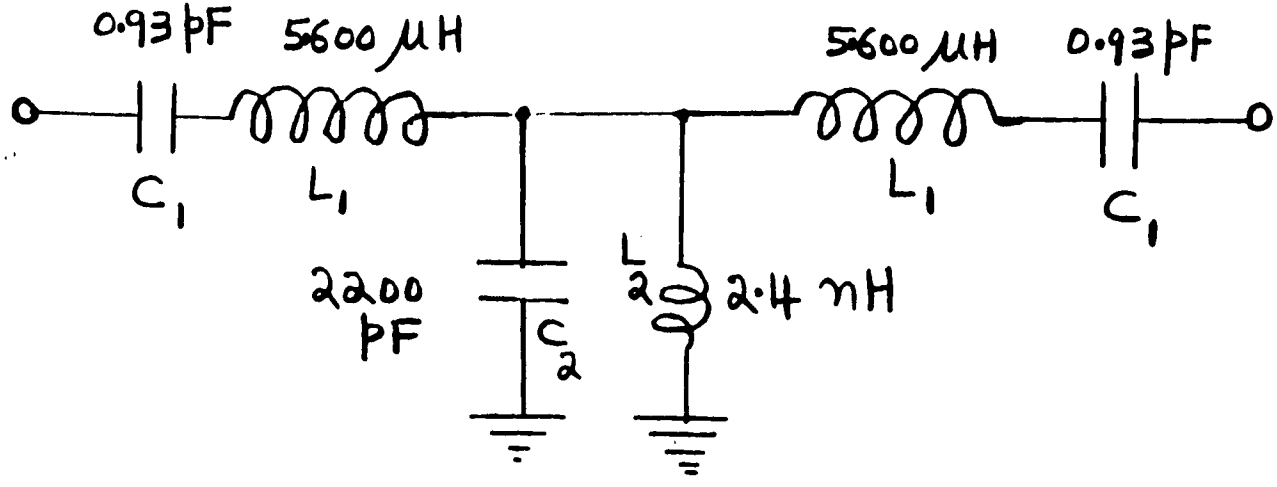


Figure 4. Band Pass Filter Circuit - The Component Values are Close to The Ones Given by ESYN

```
Touchstone (TM) - Configuration( 100 1500 100 16062 1997 1000 1 3294 )
              CB3F70B2.CKT     Tue Jul 28 14:49:39 1987

! ============================06/30/87 - 10:09:08=========================

!                         LUMPED CHEBYSHEV BANDPASS NETWORK
!                       FREQUENCY : 69.00000 to 70.50000 MHz
!     INPUT TERM   : R = 50.00000 Ohms
!     OUTPUT TERM  : R = 50.00000 Ohms
!                                    3 RESONATORS
!     0.100000 dB RIPPLE           0.000000 dB MIL          0.000000 dB SLOPE

! ============================================================================


DIM
    FREQ MHZ
    RES  OH
    IND  NH
    CAP  PF
    LNG  MIL
    ANG  DEG

CKT
    SLC      1    2   L = 5472.595   C = 0.951500
    PLC      2    0   L = 2.138599   C = 2434.852
    SLC      2    3   L = 5472.595   C = 0.951500
DEF2P        1    3   SYN

    RES      1    0   R = 50.00000
DEF1P        1        R1

    RES      1    0   R = 50.00000
DEF1P        1        R2

OUT
    SYN   DB[S21]   GR1

FREQ
    SWEEP     69.00000   70.50000   0.107143

GRID
    RANGE     69.00000   70.50000   0.150000
    GR1       0.000000   -1.00000   0.100000

TERM
    SYN       R1   R2
```

Figure 5. TOUCHSTONE Circuit File for BPF

27-12

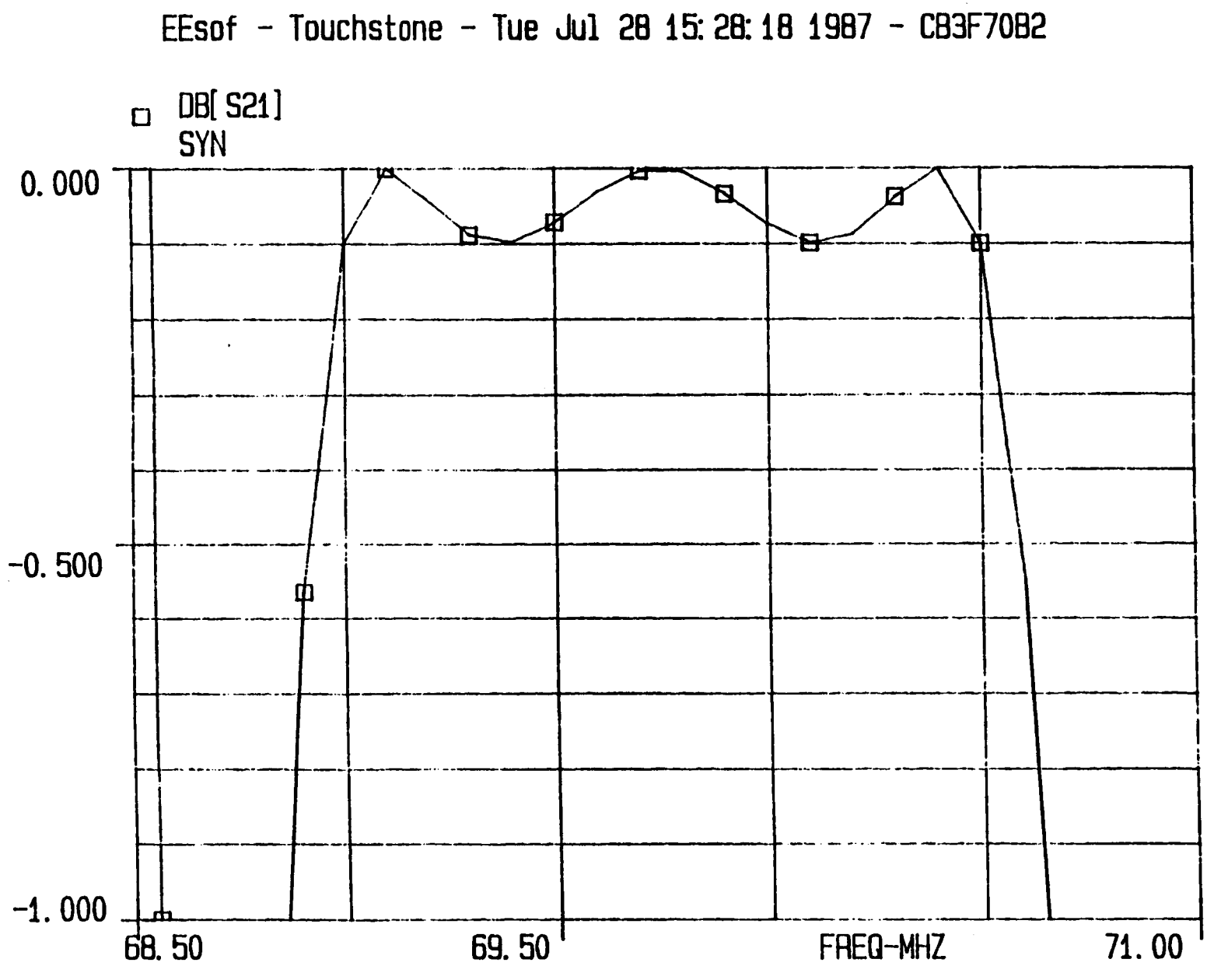


Figure 6. TOUCHSTONE Plot for BPF

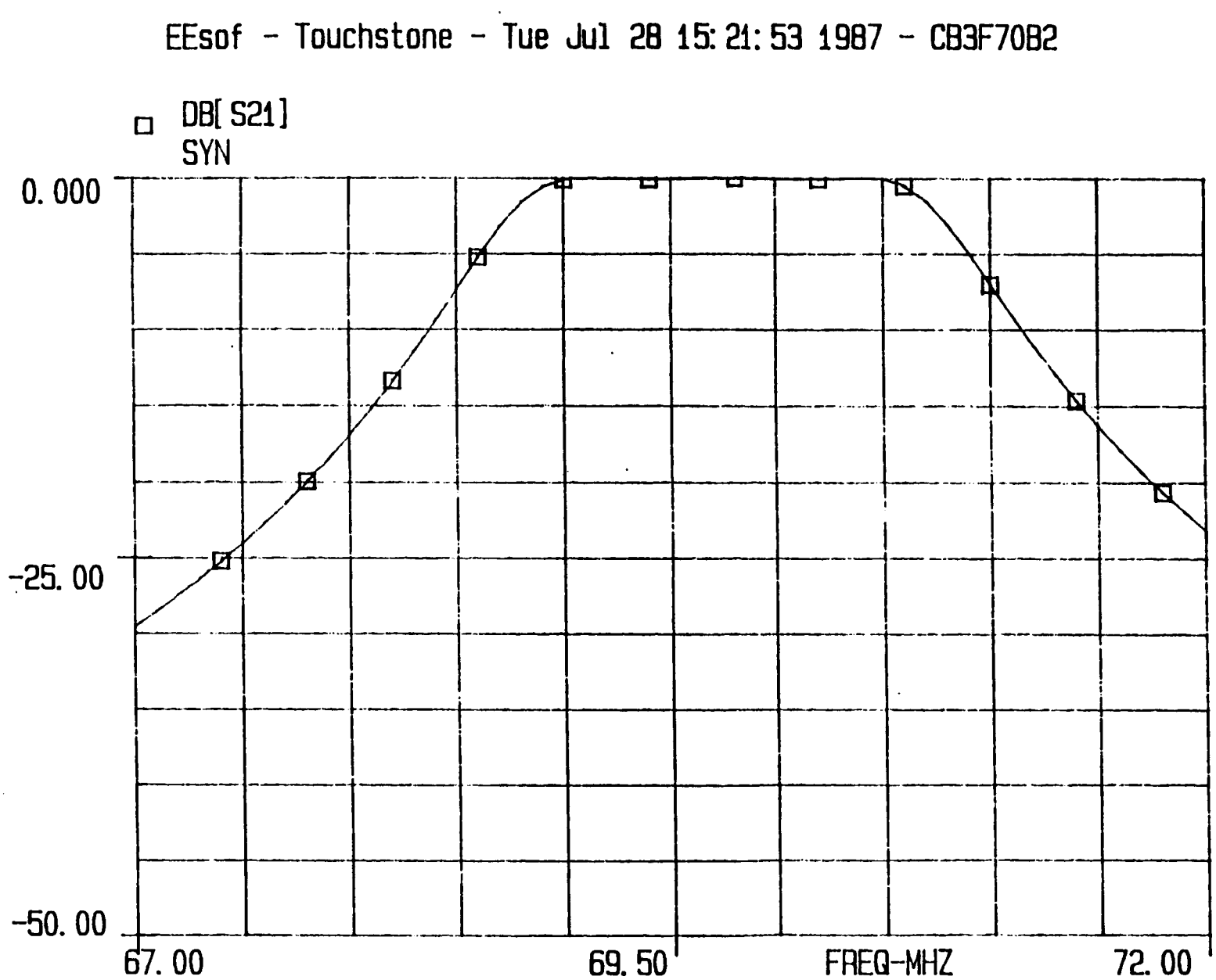


Figure 7. TOUCHSTONE Filter Plot

1100 PF value was used and the inductor value was about 5 nanohenrys. The frequency response of the filter with its housing and SMA connectors is shown in Figure 8. The return loss appears at channel 1 in Figure 8 and the pass through loss at channel 2. The filter had more insertion loss and wider bandwidth than expected. A part of the problem was getting all sections to resonant at the same frequency and having to deal with a new nanohenry inductor. The filter circuit design given by ESYN gave the starting component values, but the actual values that gave close to satisfactory performance was found by trial and error and they were:

L_1 = 1.50 - 1.59 MH

C_1 = .5 - 2.5 PF

L_2 = 5 NH

C_2 = 1100 PF

Cascaded Amplifier: This is a high gain (54 dB) low noise (NF<2.7 dB) AVANTEK UTC5-213 amplifier. The amplifier was tested using HP Sweep Generator and Scaler Network Analyser and was found to yield a gain of 56 dB with less than 0.5 dB gain flatness in the 50 - 90 MHz range.

Peak Detector: A UTD-1000 level detector was selected as it had an input impedance of 50 ohms. The detector consists of an active high to low impedance converter that drives a Schottky-barrier detector diode. Matched back-to-back silicon diodes which are closely thermally-coupled to the detector provide a dc tracking reference. The output of

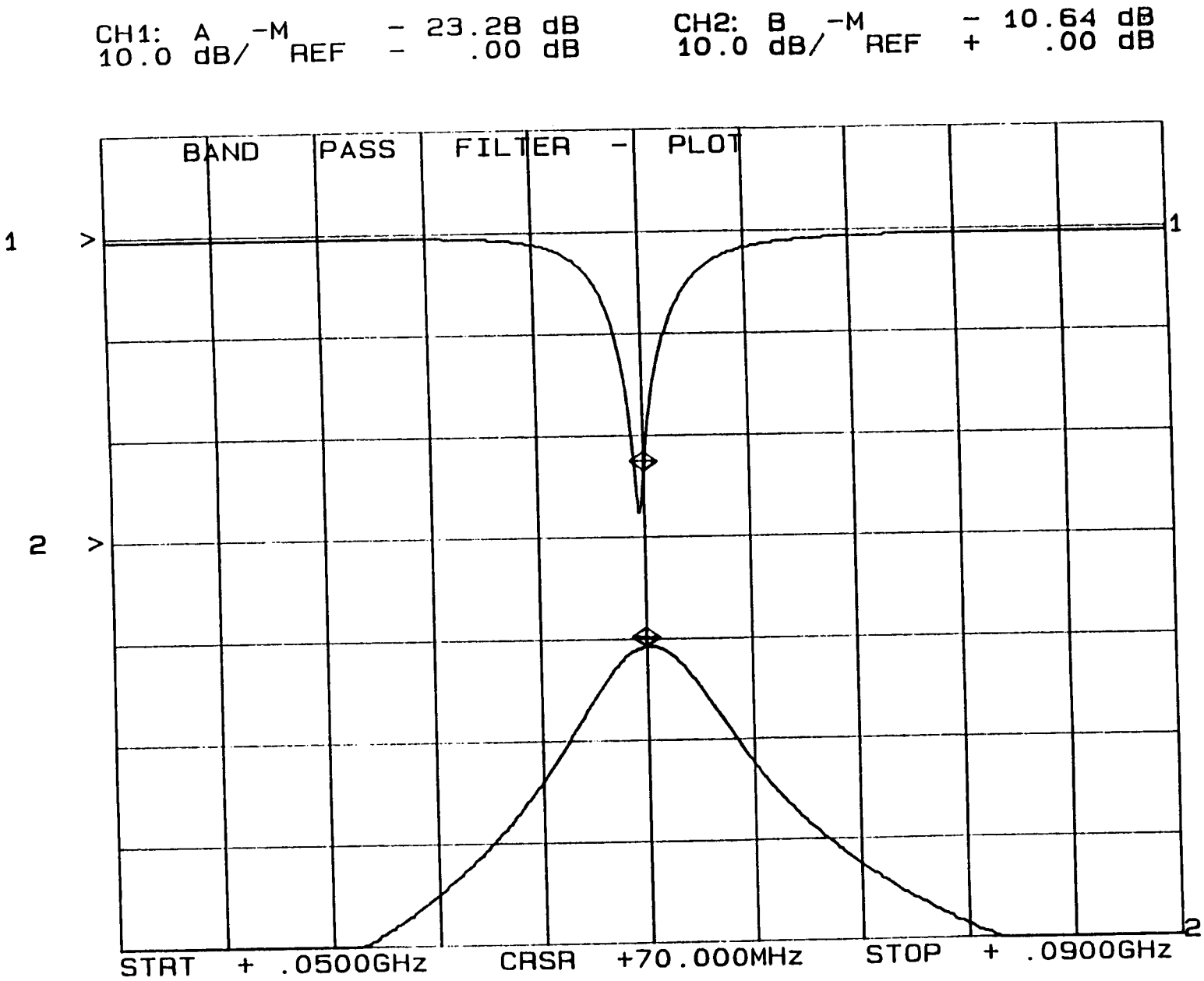


Figure 8. Filter Plot Using HP Scaler Network Analyzer

the detector is a dc voltage that varies in the - 5 mv to - 650 mv range as the input signal power varies from - 30 to + 10 dBm. An inverter amplifier with a voltage gain of 15 is built using an OPAMP 3554, which changes the output of the peak detector to the range .1 - 10V. The OPAMP has a high slew rate, a high gain-bandwidth product and fast settling time. The output of the inverter is fed to a very high speed 8 bit resolution A/D converter that has conversion time less than one microsecond. The 8 bit digital signal will be applied to a signal processor. The signal processor will be programmed to compare the digital signals from different antennas and issue a control signal to the microwave switch on the top of Figure 1 that will switch it to the antenna with the highest signal strength. Mr. John Ngo was assigned the responsibility to look into the problem of programming the signal processor to get the above mentioned task accomplished.

Directional Coupler: Four directional couplers are shown in Figure 1 placed between antennas and microwave switches. The ones purchased were 10 dB couplers made by ARRA, Model #9164-10. They had low insertion loss (<.6 dB) and high directivity (>25 dB). They were tested on the Scalar Network Analyzer.

Microwave Switches: Two microwave single-pole multi-throw diode switches that have low insertion loss and high

isolation between the output ports were ordered. They are made by Harold Tek and the model # is S2D182-32. Their delivery is being awaited presently.

Conclusions

The automatic antenna switching design for EVA communication link with Base Station as proposed by RCA was studied. The hardware needed for antenna signal monitoring was selected and ordered. As they arrived, they were tested. The bandpass filter at IF was designed, fabricated, and tested. The peak detector and the inverting amplifier that followed it were breadboarded and tested. Due to the delay in the procurement of the components needed, the assembly of signal level detector could not be completed. A request was made that someone at NASA/JSC be assigned to complete the project, test the switching system design and study the effect of switching on the transmitted data.

Acknowledgements

It is my pleasure to thank my NASA colleagues, Dr. G. D. Arndt and Mr. John Ngo of the Electromagnetic Systems Branch at NASA/JSC for their continued help and support during my summer stay. I would also like to express my thanks to Mr. Roland Shaw and Jeff Kovitz for their help during fabrication and testing of the bandpass filter. My sincere thanks to Margie Robertson for typing this report.

SPACE STATION SOFTWARE RELIABILITY ANALYSIS BASED ON FAILURES OBSERVED DURING TESTING AT THE MULTISYSTEM INTEGRATION FACILITY

Final Report

NASA/ASEE Summer Faculty Fellowship Program-1987

Johnson Space Center

Prepared by:	Tak Chai Tamayo
Academic Rank:	Assistant Professor
University Department:	University of Houston-UP Department of Industrial Engineering Houston, Texas 77004

NASA/JSC

Directorate:	Mission Support
Division:	Spacecraft Software Division
Branch:	Systems Development
JSC Colleague:	Richard E. Coblentz
Date:	August 14,1987
Contract Number:	NGT 44-001-800

ABSTRACT

Quality of software not only is vital to the success operation of the Space Station, it is also an important factor in establishing testing requirements, time needed for software verification and integration as well as launching schedules for the Space Station. Defense of management decisions can be greatly strengthened by combining engineering judgements with statistical analysis. Unlike hardware, software has the characteristics of no wearout and costly redunduncies, thus making traditional statistical analysis not suitable in evaluating reliability of software.

A statistical model was developed to provide a representation of the number as well as types of failures occur during software testing and verification. From this model, quantitative measure of software reliability based on failure history during testing are derived. Criteria to terminate testing based on reliability objectives and methods to estimate the expected number of fixings required are also presented here.

INTRODUCTION

The purpose of Multisystem Integration Facility (MSIF) is to provide a facility on which information systems for the Space Station, which are produced by different developers, may be integrated, tested, verified, certified for flight, and packaged for launch. The MSIF concept was motivated by the facts that Space Station softwares are being developed by multiple developers at different sites. The systems are highly distributed and will be built up in phases, over a number of launches. Several upgrades and changes will take place over the life of the Space Station. MSIF will be required to first perform testing using computer models of all the Space Station systems. As real systems are delivered at MSIF, testing will be performed using combinations of models and real systems. The final test will be one in which all systems are actual flight-ready versions. Since the correction of errors found during multisystem integration is the responsibility of the developer, control over delivered systems may be returned to the developer for the correction of errors, and then back to the MSIF to continue testing.

Software is an important element of the Space Station, and is vital to its successful operation. Failure of softwares can be life-threatening in some cases. In addition, the quality of software can greatly affect the amount of fixings required during the testing, integration or verification process, thus making it possible to cause delays in launching of the Space Station, which is scheduled to begin in January, 1994. Consequently there is an urgent need to search for a quantitative measure of the reliability of the software, and to develop methods of combining reliability of software and hardware elements of

the Space Station to establish the system reliability of the Station. The concept of software reliability differs from that of hardware reliability in that failure is not due to a "wearing out" process. Software failures are in fact errors which, owing to hte complexity of a computer program, do not become evident until the combination of conditions bring the error to light. Unlike the hardware bathtub curve, there is no wearout characteristics, but only a continuing burn-in. Once a software error is identified and properly fixed, it is in general, fixed for all time. However, the large number of possible paths and its inputs in a space station software makes complete testing of the software generally impossible.

Several approaches are currently available for testing of a software: path testing, functional testing and formal proofs of correctness. A complete functional test would verify that the correct output is produced for each input . It would consist of subjecting the program to all possible input streams. However, a ten-character string has 2^{80} possible input streams and corresponding outputs. So complete functional testing in this sense is clearly impractical. In path testing, one would design a sufficient number of test cases to assume that every path through the routine is exercised at least once. But most often, even the number of paths through a small routine can be astronomical to permit all paths to be tested. As for formal proofs of correctness, each program statement is examined and used in a step of an inductive proof that the routine will produce the correct output as stated by formal mathematics. The practical issue here is that such proofs are very expensive and have been applied only to numerical routines. Not only are all known approaches to absolute demonstrations of error-free

impractical, they are impossible as well.

Because exhaustively tested and error-free software are not made possible by current acceptance procedures, purchaser of a software product is provided with no quantitative information on which to base an acceptance decision and is thus forced to make these decisions based mostly on intuition and his own experience in similiar situation. Therefore our goal should be to provide sufficient testing to assure that the probability of failure due to hibernating errors is sufficiently low to be acceptable. It is expected the level of testing required will depend on the system/component, criticality and complexity, state of development and cost and usage of the system.

Software reliability is defined here as the probability that a given software operates for some time period without software error detectable by executing the codes on the machine for which it was designed, assuming that it is used within design limits. Such being the case, test cases should be designed to cover the operating scenarios of the information system designed. When softwares are delivered to MSIF, they have already been successfully tested on the flight-compatible hardware. MSIF testing will start using models of other systems, and progress to using delivered versions of the other systems. Current concepts of MSIF requires if errors are detected, the software be returned to its developer with descrepancy reports of the errors found. After proper fixing, the software is returned to MSIF for retesting. In general, while fixing the errors found, new errors are also introduced. A portion of the old errors persisted, and will reoccur during the retesting. This process is repeated until a decision is made about the quality of the software based on the testing results. In the past it normally means the

software operates successfully on all test cases it was subjected to. The goals of this paper are:

1. Develop a statistical model to describe the failures behaviour during testing.
2. Obtain a statistical measure of software reliability, based on failure history observed during testing.
3. According to prespecified software reliability and failure history, establish criteria as to when testing of software can stop.
4. Combining reliability of software and hardware elements of the Space Station to establish "system" reliability of the Station.
5. Specified types of error records to be maintained during testing so that they can be used for later statistical analysis.

These goals are motivated by the fact that in the Space Shuttle program, the extent and degree of testings performed on space softwares have generally been made based on management judements. Verification requirements are to be determined individually for each hardware/ software product, based on criticality and risk associated with the hardware/software when it is integrated into the operational enviroment.

DESCRIPTION OF THE MODEL

The model of interested here is the failure behavior of a software after it is delivered to MSIF for testing and integration. A number of test cases designed to cover a selection of the enviroment in which the software will be used are run and errors occured during execution are recorded in descrepancy reports. The software is then returned to its developer for fixing. After proper fixing, it is returned to MSIF for retesting, where a portion of the old errors may reoccur and some new errors are detected.

Assumptions of the Model

1. All errors are caused only by the faults in the software, thus all others involved in testing are assumed to possess high fidelity.
2. All errors occure during testing are observed.
3. The number of new errors found are statistically independent of the total number of errors found during the previous trial.
4. The failure rate of new errors for each trial is dependent of the number of fixings already performed on the software.
5. The number of test cases run during each trial remains relatively constant.
6. All persisted errors are statistically independent of each other.
7. The number of new errors observed during each trial follows a poisson process.

Although it is possible that failure rate of new errors found

during each trial can be directly proportional, constant, or inversely proportional to the number of fixings performed, historical information gathered during the development of the Shuttle Orbiter primary flight software indicates a decreasing trend.

During each trial, the total number of errors detected consists of two independent entities: new and persisted errors.

Let

N_k = total number of errors detected during kth trial

X_k = number of new errors detected during kth trial, after (k-1) fixings by the developer

R_k = number of persisting errors from the previous trial.

Thus, the total number of errors detected at each trial is the sum of the number of new errors introduced by the last fixing and the number of errors persisted from the last trial, i.e.

$$N_k = X_k + R_k.$$

Analysis of the Model

Let

p_k = probability of an error found in kth trial to persist in (k+1)th trial

If the number of errors found during kth trial is n_k, then the probabaility of r errors persist in (k+1)th trial after fixing is:

$$\text{Prob}(R_{k+1} = r \mid N_k = n_k) = C(n_k,r)p_k^{\,r}(1-p_k)^{n_k-r}, \qquad r = 0,1,2, \ldots, n_k$$

where

$C(n_k,r)$ = the number of unordered samples of size r taken from n_k.

This is the conditional probability density of persisting errors based on the total number of errors found in previous trial and it follows a binomial distribution $B(n_k, p_k)$. As defined in the model, the total number of errors in the next trial is determined by the sum of two independent random variables, namely the numbers of new and persisted errors. This implies future errors is dependent of the number of errors found at present through persisted errors. Therefore the conditional probability density function of future errors based on present condition is:

$$\text{Prob}(N_{k+1} = n \mid N_k = n_k) = B(n_k, p_k)^* g(X_{k+1}).$$

where

$g(X_{k+1})$ is the probability density function of the number of new errors found during (k+1)th trial, and B*g represents the convolution of the two random variables, R_{k+1} and X_{k+1}.

By the assumption that X_{k+1} follows a poisson distribution with mean λ_{k+1}, the convolution of a binomial and poisson distributions is given as follows:

$$\text{Prob}(N_{k+1} = n \mid N_k = n) = \sum_{j=0}^{\min(n,n_k)} C(n_k, j) p_k^{\,j} (1-p_k)^{n_k - j} \exp(-\lambda_{k+1})(\lambda_{k+1})^{n-j}/j!$$

If p_k is relatively small, then the density function of $B(n_k, p_k)$ can be approximated with a poisson distribution, and the convolution of R_{k+1} and X_{k+1} is a poisson process given as:

$$\text{Prob}(N_{k+1} = n \mid N_k = n_k) \approx \exp(-\lambda_{k+1} - \mu_{k+1})\,(\lambda_{k+1} + \mu_{k+1})^n/n!$$

$$\text{with mean} = \lambda_{k+1} + \mu_{k+1} \quad \text{and} \quad \mu_{k+1} = n_k p_k.$$

In the case where n = 0, this conditional density function estimates the software reliability based on number of failures observed.

Criteria for Termination of Testing:

Extent of testing for softwares shall be conducted at a level consistant with its criticality level associated with the Space Station. Softwares that are highly critical to the successful operation of the Station will require high reliability, thus more detail testing than the others. Current concept documents of MSIF states thaat the degree to which a system is tested at MSIF depends on its risk category and how tightly coupled it is with the Data Management Systems (DMS). Suppose a particular software is required to have a minimum reliability level,

say REL, during the mission period T. In particular, REL is defined as the probability no error will occur during the mission period, and (1-REL) is the probability any error is detected during T. Hence the following inequality must be satisfied in order to meet the required reliability level REL.

$$REL \geq Prob(N_{k+1} = 0 \mid N_k = n_k)$$

Using the approximation of a poisson process, it is thus suffucient to solve for n_k^* such that:

$$REL \geq \exp(-\lambda_{k+1} - n_k p_k)$$

or by taking logorithm on both sides,

$$n_k^* \geq [\, -\ln(REL) - \lambda_{k+1} \,] / p_k \qquad (1)$$

Equation (1) gives the set (k, n_k^*) which yields the required reliability level REL. Since $n_k^* \geq 0$, it implies that the earliest time testing may terminate can be obtained by setting $n_k^* = 0$ and then solve for a smallest k that satisfies:

$$\lambda_{k+1} \leq -\ln(REL).$$

Let m be the expected number of fixings required before a software passes the testing and

$$m = E[k]$$

$$= \sum_{k=0}^{\infty} k \, \text{Prob(testing is terminated after k fixings)}$$

$$= \sum_{k=0}^{\infty} k \, \text{Prob}(N_k = n_k^*, n_k^* \geq 0)$$

which can be obtained by applying properties of conditional probabilities as follows:

$$\text{Prob}(N_k = n_k^*, n_k^* \geq 0)$$

$$= \sum_{n_{k-1}=0}^{\infty} \text{Prob}(N_k = n_k^*, n_k^* \geq 0, N_{k-1} = n_{k-1})$$

$$= \sum_{n_{k-1}=0}^{\infty} \text{Prob}(N_k = n_k^*, n_k^* \geq 0 \mid N_{k-1} = n_{k-1}) \, \text{Prob}(N_{k-1} = n_{k-1})$$

and

$$\text{Prob}(N_j = n_j) = \sum_{n_{j-1}=0}^{\infty} \text{Prob}(N_j = n_j \mid N_{j-1} = n_{j-1}) \, \text{Prob}(N_{j-1} = n_{j-1}) \, , \quad .$$

$$\text{for } j = 1, 2, \ldots, n_{k-1}.$$

Example

Suppose from historical data, one will observe X_k new errors after k fixings and X_k follows a poisson process with mean λ_k = 0.5exp(-k) per KSLOC[1] per hour in execution time. Ten percent of the errors found during a trial are corrected by the developer after one fixing. One hundred test cases which takes a total of 100 hours to execute are designed to test the Atmosphere Control and Supply (ACS) subsystem software, which provides total and partial pressure control within the pressurized habitation module in the Space Station. Such system consists of 10 KSLOC and is classified as criticality level 1 which requires a minimum reliability level of 0.999 during its useful life cycle of 20 years, i.e. the probability of no error being detected during the next 20 years of operation is 0.999.

By adjusting the unit of measurement, the mean number of failures for the ACS software during each trial period which lasts 100 hours is:

$$\lambda_k = 50\exp(-k)$$

The number of hours in 20 years = 175,200 and is equivalent to 1,752 trial periods.

To achieve reliability of 0.999 during the next 20 years, the software should achieve a minimum reliability level REL^* during the test period of 100 hours where

$$(REL^*)^{1752} = 0.999$$

or $$REL = 0.9999994$$

[1] KSLOC = thousand source line of codes

The minimum number of fixings needed for the ACS software based on the required relaibility level is given in Table 1. It shows that for required reliability of 0.9999994, the earliest time testing can terminate is after the 18th fixing when no error is detected during that trial. An estimate of the software reliability during the next period based on selected number of errors observed during testing are given in Table 2.

Table 1. The Minimum Number of Fixings Needed for the ACS Software Based on Required Reliability Level

Reliability Level (REL)	Minimum Number of Fixings (k)
0.90	6
0.95	6
0.99	8
0.999	10
0.9999	13
0.99999	15
0.999999	17
0.9999997	18
0.9999999	19

Table 2. Software Reliability Based on Number of Errors Observed

Trial (k)	# Errors Observed During Testing (N_k)	Reliability
1	0	0.0011514
1	1	0.0010418
2	0	0.0829635
2	1	0.0750685
3	0	0.4002035
3	1	0.3621191
4	0	0.7139821
4	1	0.6460377
5	0	0.8834349
5	1	0.7993650
6	0	0.9554297
6	1	0.8645085
7	0	0.9833667
7	1	0.8897870
8	0	0.9938485
8	1	0.8992713
9	0	0.9977325
9	1	0.9027858
9	2	0.8168743
10	0	0.9991652
10	1	0.9040821
10	2	0.8180473
11	0	0.9996928
11	1	0.9045595
11	2	0.8184792
12	0	0.9998869
12	1	0.9047351
12	2	0.8186382
13	0	0.9999584
13	1	0.9047998
13	2	0.8186967
14	0	0.9999847
15	0	0.9999943
16	0	0.9999979
16	1	0.9048355
17	0	0.9999992
18	0	0.9999997
19	0	0.9999999

C-3

Collection of Data

Certain data concerning the software will have to be collected in order to verify the statistical model described here. This includes:

1. The frequency of persisted errors from previous trial and new errors that are introduced during the fixing effort.
2. The number of fixings performed by the developer when these errors were found.
3. Criticality level of errors found at each trial.
4. A function which relates the failure rates of new errors at each trial with the number of fixings performed.
5. Probability distributions of the number of new errors detected at each trial.
6. Probability an identified error is corrected by the developer through one fixing.
7. Number of test cases applied on each trial.
8. Size of the software.

The function which relates failure rates of new errors with the number of fixings can be obtained by applying regression analysis on the frequency of failures obtained through historical data. Since a perfect fit of n experimental data may require a polynomial of degree (n-1), techniques of selecting a "sufficient" function may be needed to reduce this polynomial to an acceptable form. If distributions of failures are unknown, a Chi-square goodness-of-fit test can be employed to test for pattern of distributions. Data from the Space Shuttle softwares were not applied to this model because they made no distinction between persisted and new errors. However, the data did support a decreasing

trend with the number of fixings. When available, the amount of time and cost associated with each fixing can be combined with the amount of computer execution times required during testing to establish testing schedules and total cost of softwar testing so that launch schedules and budgets are met. Factor like degree of interaction with other systems is important in determining the reliability of a distributed system, and therefore should also be included in the data when available.

System Reliability

The Challenger accident shows that in case of accidents, defense of management decisions can be greatly strengthened if they are made based on combination of statistical analysis and engineering judgements. In the context of manned Space Station, it is important to explore reliability theory to assess the risk of extended human presence in space. Most Space Station systems are complex systems composed of hardware and software, both of which are required to be in operational states in order for the system to perform its designed function. It is thus necessary to include software as part of the components which form the reliability network of the Space Station. It has been shown that a system with subsystems and components in series will have reliability less than that of its weakest link. Suppose the ACS subsystem hardware in the habitation module has failure times which follows an exponential distribution with MTBF = 100 years and reliability of its software which interacts with the Data Management Systems (DMS) is 0.999 during the next 20 years. Then the ACS

subsystem will have reliability less than those of the hardware, software and DMS alone. In particular, if no scheduled maintenance work are to be performed during the next 20 years and the DMS has reliability of 0.995, then the reliability of the ACS subsystem is:

$$\begin{aligned} \text{Reliability} &= \exp(-20/100) \times 0.999 \times 0.995 \\ &= 0.8138224. \end{aligned}$$

This demonstrates an important fact that the reliability of a complex system decreases rapidly as more subsystems are added to the system design. For instance, a system which requires five components in series configuration will have reliability of 0.77 if each of the component was tested to have reliability of 0.95. Thus in order to achieve the goal of a highly reliable system, efforts should be made to obtain highly reliable hardware as well as software through either engineering design or testing. While traditional methods of redunduncy works well with hardware, it can be quite costly for complex software, as redunduncy in software normally means an independent development of the computer program. It is this unique characteristic of software which makes software testing an effective method to maintain software quality.

CONCLUSION

The lack of quantitative method to evaluate reliability of software delivered by the developer motivated the statistical model designed here. The unique characteristics of no wearout and costly redundancy has made software testing an only way besides software design to maintain software quality. The model developed here represents the failure pattern during software testing, which includes new errors introduced by the fixing and persisted errors from previous trial. Quantitative approaches were derived to predict the software reliability and criteria to terminate testing based on failure history. These results can be applied to enhance the safety of the Space Station and to avoid delays in launch schedules due to delay in the software verification process.

REFERENCES

1. Preliminary Space Station Multisystem Integration Facility Document, NASA JSC, June, 1987.

2. Musa, J. D., "A theory of Software Reliability and its Applications," IEEE Transaction on Software Engineering, Vol. SE-1, No. 3, pp. 312-327, September, 1975.

3. Musa, J. D., "Software Reliability Measurement," The Journal of Systems and Software, Vol. 1, pp. 223-241, Elsevier North Holland, Inc., 1980.

4. Final Report for "The Analysis of the Statistical and Historical Information Gathered During the Development of the Shuttle Orbiter Primary Flight Software," N82-29960, Sumitted by Texas A&M Research Foundation.

5. Feller, William, An Introduction to Probability Theory and its Applications, John Wiley & Sons, Inc.

6. Smith, David J., Reliability and Maintainability in Perspective, Practical, Contractual, Commercial and Software Aspects, Second Edition, John Wiley & sons, Inc.

7. Architectual Control Document for Enviromental Control and Life Support System, NASA JSC 30262, January, 1987.

ANALYSIS AND INTERPRETATION OF SATELLITE FRAGMENTATION DATA

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1987

Johnson Space Center

Prepared by:	Arjun Tan
Academic Rank:	Associate Professor
University & Department:	Alabama A&M University Department of Physics Normal, Alabama 35762
NASA/JSC	
Directorate:	Space and Life Sciences Directorate
Division:	Solar System Exploration Division
Branch:	Space Science Branch
JSC Colleague:	Gautam D. Badhwar
Date:	August 14, 1987
Contract Number	NGT 44-001-800

ABSTRACT

The velocity perturbations of the fragments of a satellite can shed valuable information regarding the nature and intensity of the fragmentation. A feasibility study on calculating the velocity perturbations from existing equations was carried out by analyzing 23 major documented fragmentation events. It was found that whereas the calculated values of the radial components of the velocity change were often unusually high, those in the two other orthogonal directions were mostly reasonable. Since the uncertainties in the radial component necessarily translate into uncertainties in the total velocity change, it is suggested that alternative expressions for the radial component of velocity be sought for the purpose of determining the cause of the fragmentation from the total velocity change. The calculated variances in the velocity perturbations in the two directions orthogonal to the radial vector indicate that they have the smallest values for collision induced breakups and the largest values for low-intensity explosion induced breakups. The corresponding variances for high-intensity explosion induced breakups generally have values intermediate between those of the two extreme categories. A three-dimensional plot of the variances in the two orthogonal velocity perturbations and the plane change angle shows a clear separation between the three major types of breakups. This information is used to reclassify a number of satellite fragmentaion events of unknown category.

INTRODUCTION

Since the launch of Sputnik 1, over 16,000 pieces of man-made objects have circled the earth (NSSC Satellite Catalog, 1985). Of these, almost 6,000 are still in orbit, but only 5 percent are useful functioning satellites (Johnson, 1985). The number of smaller, untrackable and uncataloged objects is believed to be several time higher (Sehnal, 1984; Taff, et al., 1984).

Almost half of the cataloged orbital "debris" originated from the fragmentation of satellites. As of January 1986, over 90 documented cases of satellite fragmentation events have taken place (Johnson, et al., 1986). Each event can create up to several hundreds of additional trackable debris plus an unknown number of untrackable ones, thus increasing the population of unwanted, hazardous material in space.

Satellites can break up from explosions or through collisions with other objects. Of these, explosion has been the major cause of breakups. According to Bess (1975), two classes of explosions are recognized. In a high-intensity explosion, the explosive charge is in direct contact with the spacecraft structure, e.g., ignition of excess fuel. In a low-intensity explosion, on the other hand, the charge is not in direct contact with the spacecraft structure, e.g., a pressure vessel explosion.

Satellites can also break up from collision with a man-made object or a natural meteorite. A collision can be part of an anti-satellite testing. In low earth orbit space, the average relative velocity of collisions between two objects is calculated to be about 10 km/s (Kessler and Cour-Palais, 1978), thus placing the collision in the hypervelocity range. The average relative angle between two colliding objects is estimated to be around 90 degrees (Chobotov, 1983).

The cataloged satellite fragmentation events have been categorized into the following three classes (Johnson, et al., 1986). About one-sixth are

due to propulsion related causes -- valve failure in a rocket motor, mixing of volatile fuels, ignition of excess fuel, etc. About a third have been deliberate -- anti-satellite testing, destruction of sensitive hardware or malfunctioning equipment. The remaining half of all breakups are of unknown cause -- there was little or no fuel on board and no trackable objects nearby. This raises the serious concern about the possibility that some of the unknown breakup cases might have been due to collision with a small untrackable object.

An early effort in using debris characteristics as a tool to distinguish between explosion and collision induced satellite breakups was carried out by Culp and McKnight (Culp, 1986). They developed a Satellite Fragmentation Event (SAFE) Test, which generates a score as to the more probable cause of fragmentation: explosion or collision. The mass distribution of the fragments as inferred from radar cross-section measurements were compared with those of earth-based explosion and collision experiments from Bess (1975). Also considered in the SAFE Test were the relative dispersion of larger fragments, orderliness and symmetry of Gabbard diagrams and the dependence of velocity perturbations on mass.

The SAFE Test has been deemed fairly successful in distinguishing between explosions and small particle collisions but is somewhat subjective in nature as it depends upon the skill of the performer. Also, the Gabbard diagrams are altered by the "age" of the data due to atmospheric drag and the best data just after the event are generally unavailable.

One parameter that is least affected by the age of the data is the plane change angle of the fragments, since the inclination of a satellite is unaffected by atmospheric drag. Badhwar, et al. (1987) utilized this fact to develop a scheme of determining the cause of satellite fragmentation which is both objective and free from the "ageing" effect of the data. Based on the distribution characteristics of the plane

change angles and radar cross-sections of known fragmentation cases, they came up with a scheme to distinguish between three classes of satellite breakups: collision induced, high-intensity explosion induced and low-intensity explosion induced breakups.

A set of quantities which can be readily calculated from the fragment data using existing equations and which can provide valuable information regarding the breakup are the velocity changes imparted to the fragments during the fragmentation event. McKnight (Culp, 1986) and Kling (1986) pursued this course to some extent, but lately, such efforts have been quietly abandoned because the results obtained were often thought to be unrealistic (Reynolds, 1987).

In the present study, we have re-examined the feasibilty of the above approach by studying several documented cases of fragmentation events and analyzing the calculated velocity perturbations by studying their frequency distributions, standard deviation and variance, and their relations to the radar cross-sections and plane change angles. New signs for characteristic features of the various types of breakups have been sought and the results of the study are then used to reclassify a few unknown categories of satellite breakups.

EQUATIONS

The changes in the three orthogonal components of velocity of a fragment Δv_r, Δv_θ and Δv_z (henceforth denoted by Δv_x) can be determined from the observed changes in the semi-major axis, eccentricity and inclination Δa, Δe and Δi from the following equations (Meirovitch, 1970):

$$\Delta a = \frac{2(1-e^2)^{\frac{1}{2}}}{n} \left(\frac{e \sin\theta}{1-e^2} \Delta v_r + \frac{a}{r} \Delta v_\theta \right) , \tag{1}$$

$$\Delta e = \frac{(1-e^2)^{\frac{1}{2}}}{na} \left(\sin\theta \, \Delta v_r + \frac{e}{ar} (1-e^2 - \frac{r^2}{a^2}) \Delta v_\theta \right) , \tag{2}$$

and $$\Delta i = \frac{r \cos(\omega+\theta)}{na^2(1-e^2)^{\frac{1}{2}}} \Delta v_x , \tag{3}$$

where a is the semi-major axis, e the eccentricity, r the radial distance, n the mean motion, ω the argument of perigee and θ the true anomaly at the point of fragmentation. Eliminating Δv_r and Δv_θ from Eqs.(1) and (2) and rewriting,

$$\Delta v_r = \frac{na^2(1-e^2)^{\frac{1}{2}}}{2er^2 \sin\theta} \left(2ae \, \Delta e - (1-e^2 - \frac{r^2}{a^2}) \Delta a \right) , \tag{4}$$

$$\Delta v_\theta = \frac{na(1-e^2)^{\frac{1}{2}}}{2r} \left(\Delta a - \frac{2ae}{1-e^2} \Delta e \right) , \tag{5}$$

and $$\Delta v_x = \frac{na^2(1-e^2)^{\frac{1}{2}}}{r \cos(\omega+\theta)} \Delta i . \tag{6}$$

The orbital elements of cataloged satellite fragmentation events are readily available from a technical report by Johnson, et al. (1986) in which the following quantities are cataloged whenever available: the inclination i, altitude of breakup h, altitude of perigee h_p, altitude of apogee h_a, true anomaly θ, the time period T, latitude λ and longitude ϕ. If r_o is the radius of the earth, we have immediately:

$$a = \frac{r_a + r_p}{2} = \frac{2r_o + h_a + h_p}{2} \quad , \tag{7}$$

$$e = \frac{r_a - r_p}{r_a + r_p} = \frac{h_a - h_p}{2r_o + h_a + h_p} \quad , \tag{8}$$

and $$r = r_o + h \quad . \tag{9}$$

Also, $$n = \frac{2\pi}{T} = \left(\frac{\mu}{a^3} \right)^{\frac{1}{2}} \quad , \tag{10}$$

where μ, the product of the universal gravitational constant G and the mass of the earth M, is referred to as the gravitational parameter. The argument of latitude $u = \omega + \theta$ can be calculated as follows. Since u is the angle between the ascending node (λ_o, ϕ_o) and the event point (λ, ϕ), we have (cf. Gellert, et al., 1977)

$$\cos u = \sin\lambda \sin\lambda_o + \cos\lambda \cos\lambda_o \cos(\phi - \phi_o) \quad . \tag{11}$$

Now $\lambda_o = 0°$ and ϕ_o can be read from the satellite map (Johnson, et al., 1986), whence

$$u = \cos^{-1} [\cos\lambda \cos(\phi - \phi_o)] \quad . \tag{12}$$

The semi-major axis, eccentricity and inclination of the fragemnts are read from NORAD data files, from which Δa, Δe and Δi are calculated.

RESULTS

Altogether 23 major satellite fragmentation events belonging to the following categories were studied: collision induced breakups, e.g., Solwind P-78; low-intensity explosion induced breakups, e.g., Delta 2nd stage rocket explosions due to hypergolic fuel ignition; high-intensity explosion induced breakups, e.g., planned fragmentations of Cosmos satellites as part of alleged Soviet ASAT tests; and unknown categories. These events contributed to nearly 80% of all documented debris from satellite fragmentation events. Double and multiple events were not considered in this study.

An inspection of Eq.(4) reveals that the values of Δv_r become exceedingly high in the following cases. First, if the fragmentation took place near the apogee (e.g., NOAA-3, NOAA-4 and Transit 4A rockets) or near perigee (e.g., Cosmos 252 and Cosmos 375 satellites), then the small denominator in Eq.(4) yields unusually high values for Δv_r. Taking natural logarithms and differentiating both sides of Eq.(4), we get

$$\frac{d(\Delta v_r)}{\Delta v_r} = \frac{dn}{n} + \frac{2\,da}{a} + \frac{\frac{1}{2}\,d(1-e^2)}{1-e^2} + \frac{df}{f} - \frac{de}{e} - \frac{2\,dr}{r} - \frac{d(\sin\theta)}{\sin\theta} \quad ,$$

where $f(a,e,\Delta a,\Delta e) = 2ae\,\Delta e - (1-e^2-\frac{r^2}{a^2})\,\Delta a$.

Simplifying and rearranging,

$$\frac{d(\Delta v_r)}{\Delta v_r} = \frac{dn}{n} + \frac{2\,da}{a} - \frac{de}{e(1-e^2)} + \frac{df}{f} - \frac{2\,dr}{r} - \cot\theta\,d\theta \quad . \qquad (13)$$

Evidently, the percentage error in Δv_r due to the last term in Eq.(13) alone is infinite at apogee ($\theta = 180°$) or perigee ($\theta = 0°$). Even with an accuracy of $d\theta = 1°$, the percentage errors owing to the last term at 1°, 2° and 3° on either side of apogee and perigee are 100%, 50% and 33% respectively. Secondly, for nearly circular orbits having eccentricities of 0.002 or less (e.g., NOAA-3, NOAA-4, NOAA-5 Landsat 3 and OPS 7613

rockets, Cosmos 1375 and Solwind satellites and Spot 1 Ariane 3rd stage rocket), the second term in Eq.(4) again dictates high values of the same sign for Δv_r. Equation (6) also indicates that Δv_x can have abnormally high values when the argument of latitude is close to 90° or 270°. However, we did not encounter an actual case like that in our examples. The values of Δv_θ, on the other hand, were always realistic.

Thus, to sum up, the calculated values of Δv_r were frequently unreliable, whereas the values of Δv_θ and Δv_x were mostly reliable. Hence the rest of this study was focussed on Δv_θ and Δv_x and not on Δv_r. Further, since the large values of Δv_r necessarily translate into large values of Δv, the net velocity change, its role in this study was also minimized.

Figure 1 is an illustrative example of the frequency distributions of the velocity perturbations. The fragments belonged to Solwind which was intentionally destroyed through hypervelocity impact as part of an alleged US ASAT Test (Kling, 1986). The distributions of Δv_θ and Δv_x were closer to Gaussian ones than that of Δv_r. There were slightly more negative values in Δv_θ and slightly more positive values in Δv_x but the Δv_r's were mostly negative. This is almost certainly due to the nearly circular orbit of Solwind prior to the breakup (e = 0.0022) as discussed earlier. The distribution of Δv is more or less exponential as expected.

The magnitude of Δv has been used as one of the prime criteria in distinguishing between explosion and collision induced breakups by McKnight (Culp, 1986). According to earth-based experiments, the larger fragments will acquire much larger velocity perturbations in an explosion than in a collision (Bess, 1975). Our calculations give the lowest values to Δv's for Cosmos 839 and Cosmos 1375, which would almost certainly place them in the collision category, in general agreement with the analysis of McKnight (Culp, 1986). The Cosmos 1275 and Solwind P-78 fragments had significanlty higher Δv's but they were still considerably smaller than those of explosive fragments. It may be noted that Cosmos 1275 is classified in the collision category by McKnight (Culp, 1986),

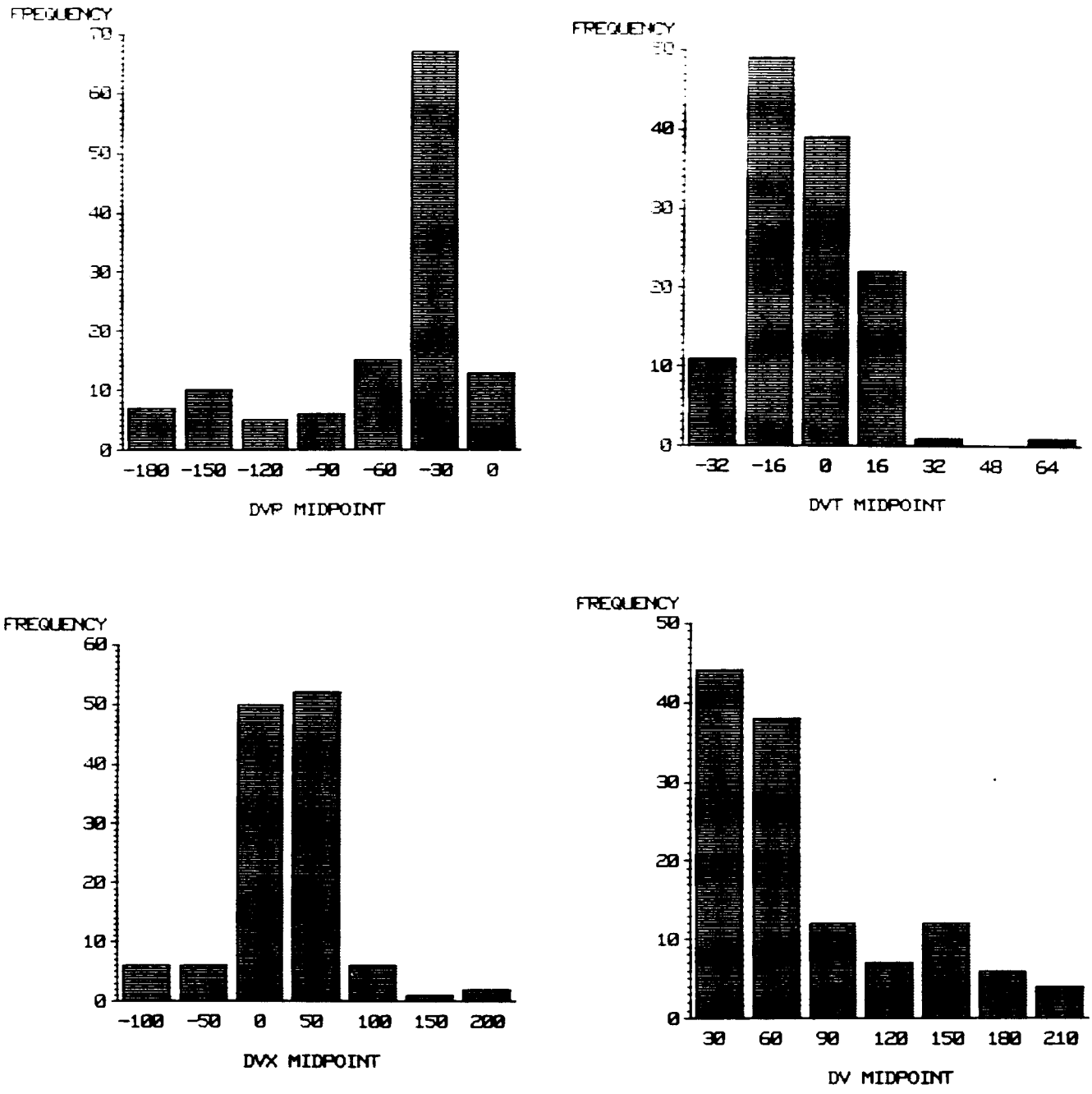


Fig. 1. Frequency distributions of Δv_r, Δv_θ, Δv_x and Δv of fragments of Solwind P-78. The velocities are in units of m/s.

but the analysis of Badhwar, et al. (1987) places a much smaller probability for that likelihood. However, one cannot put too much emphasis on the magnitudes of the velocity perturbations bearing in mind the uncertainties in Δv_r and consequently those in Δv.

Figure 2 is a two-dimensional plot of the velocity perturbations of Solwind fragments in the θ-x plane. The relative mgnitudes of the rcs values of the fragments are also indicated in the diagram. In the case of a collision with a small object, one would expect the smaller fragments to be dispersed more than the larger ones (Bess, 1975), but that is not very evident from the diagram. One must recall, however, that Solwind was made to collide with an object of comparable mass (Kling, 1986). Moreover, a collision can range from a head-on collision to a glancing one and each can have its own characteristics.

Figure 3 is a three-dimensional plot of the orthogonal components of the velocity perturbations of Solwind fragments. The relative magnitudes of the rcs values are, once again indicated in the diagram. There seems to be the appearence of a main cluster of fragments together with a more dispersed secondary one. But in view of the uncertainties in Δv_r, this feature cannot be taken seriously.

Figure 4 is a plot of the frequency distributions of velocity perturbations of the Solwind fragments in the θ-x plane. A close scrutiny reveals the appearence of a relatively small scatter of the fragments. Also conspicuous are two well-defined peaks, which was a unique feature among all the fragmentation cases studied. Whether the two peaks belonged to the interceptor and the target parents is impossible to ascertain at present.

Figure 5 is a similar plot for the fragments of NOAA-5 second stage rocket, which is one of the several cases of low-intensity explosion involving a Delta second stage rocket. Here, the scatter of the fragments is relatively large. Also there is only one main peak in the frequency plot.

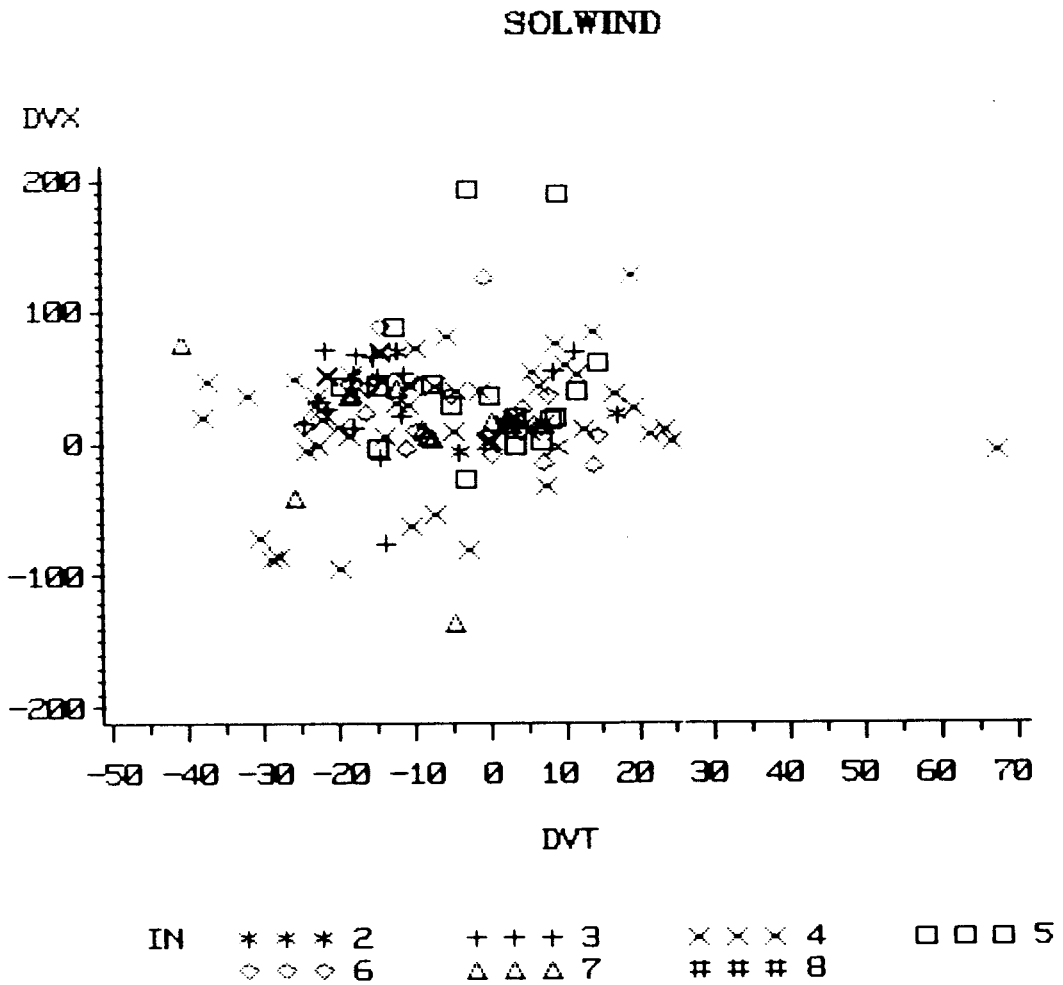


Fig. 2. Velocity perturbations plot of Solwind fragments in the θ - x plane. Also shown in the figure are the rcs values of the fragments in increasing numerical order.

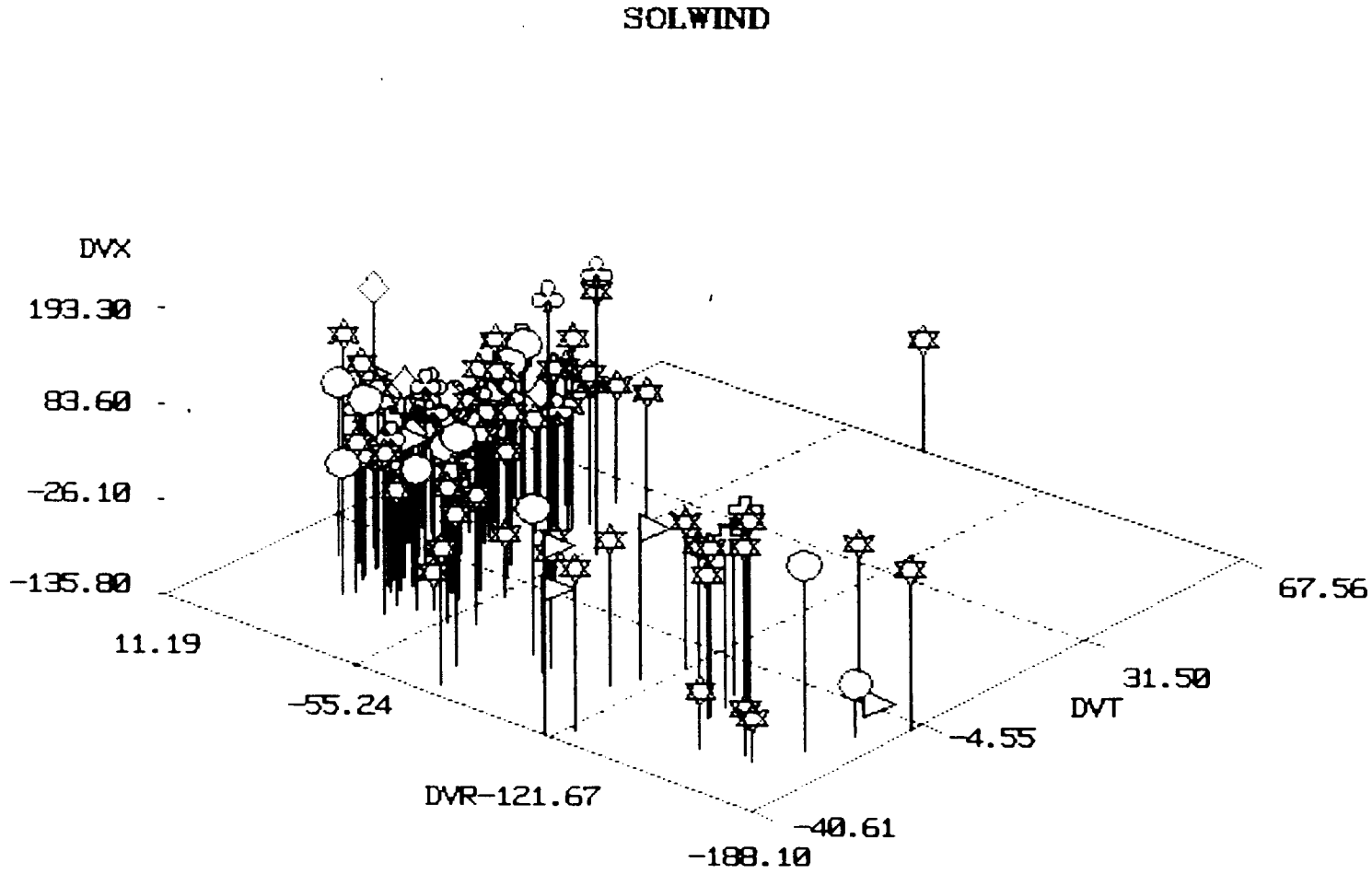


Fig. 3. Three-dimensional plot of velocity perturbations of Solwind fragments. The shapes indicate the magnitudes of rcs in the following increasing order: cylinder, cross, balloon, star, club, diamond, flag, heart, pyramid and square.

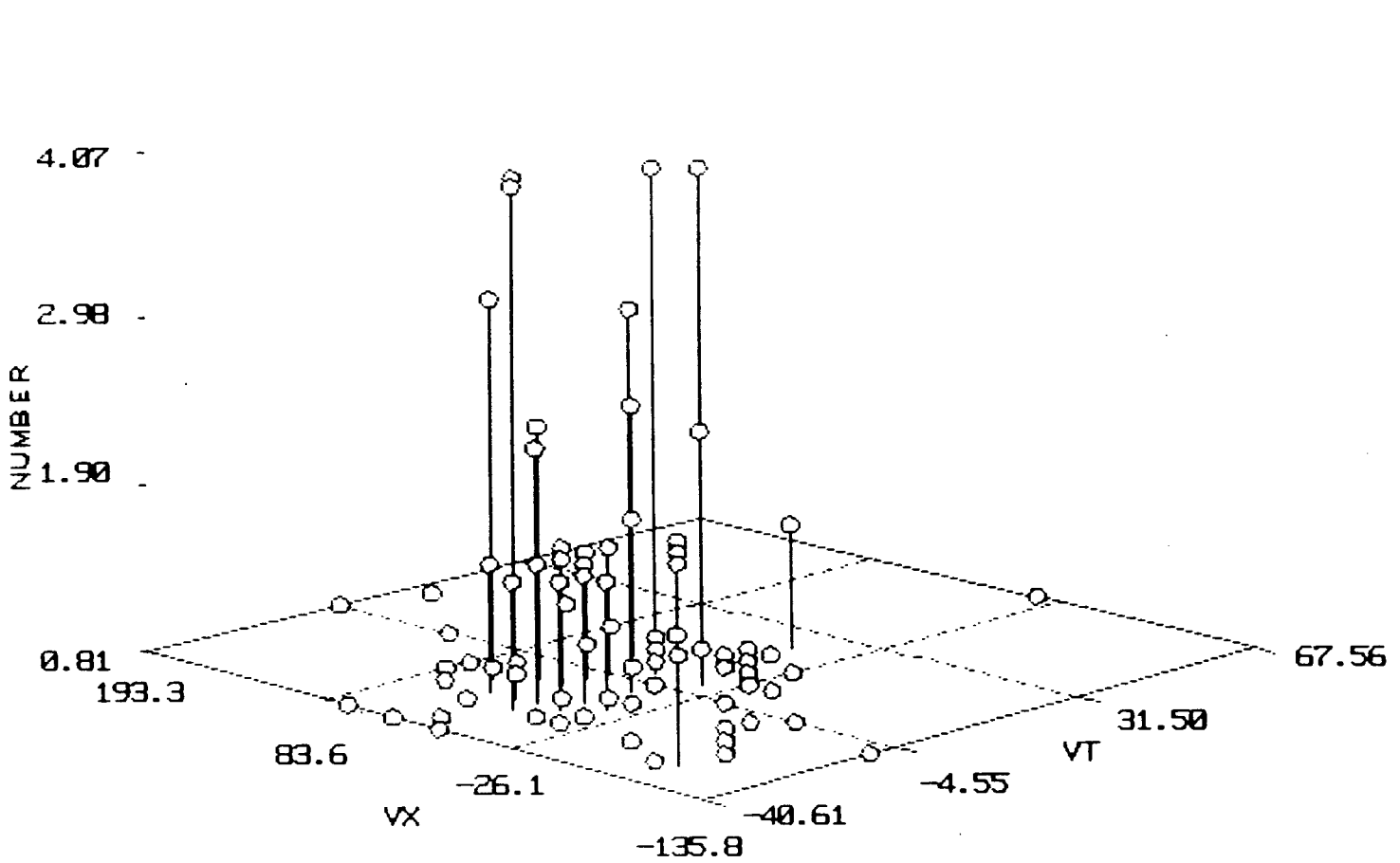


Fig. 4. Frequency distrbution of velocity perturbations of the Solwind fragments in the θ-x plane. The velocities are in the units of m/s.

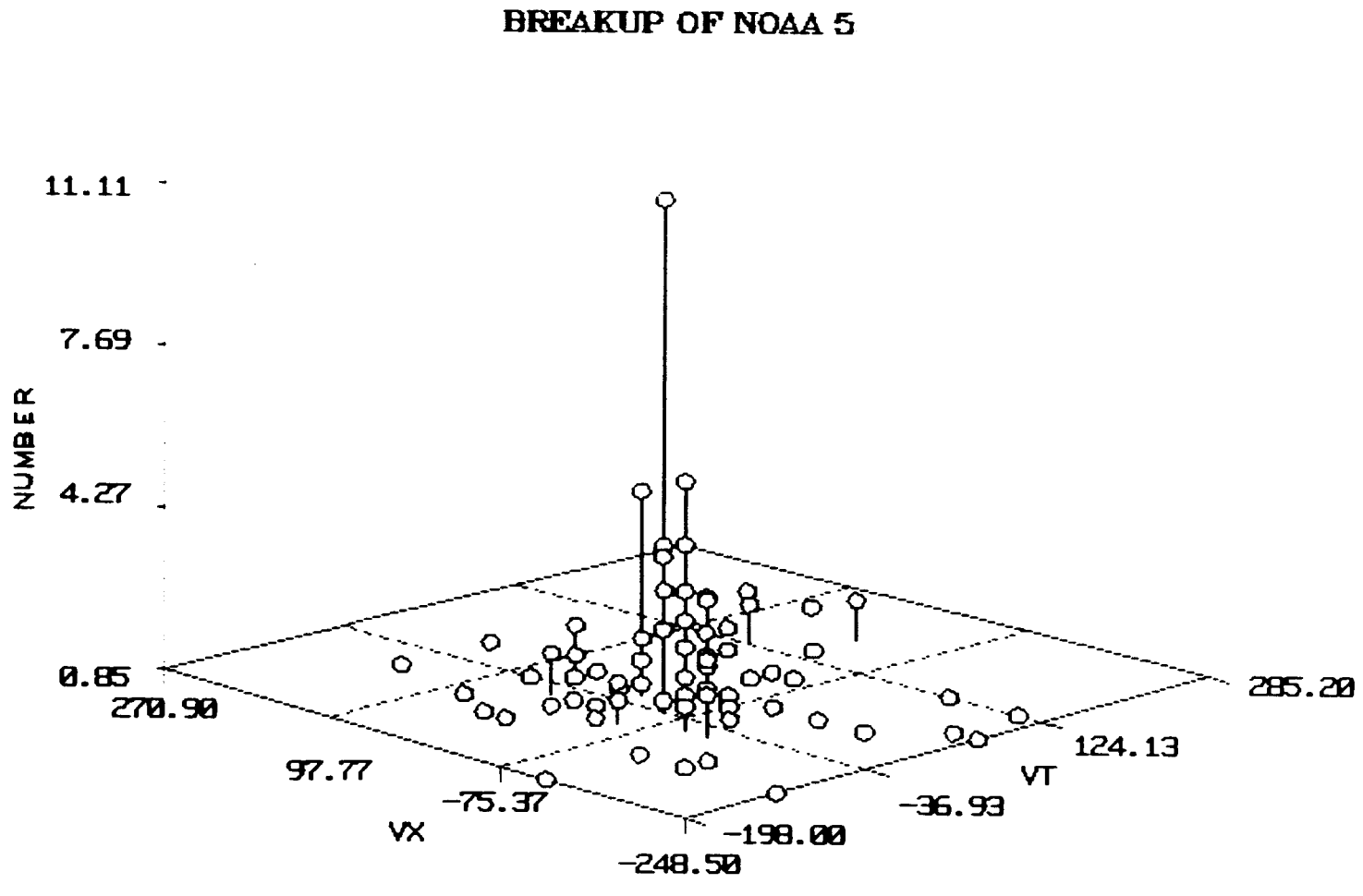


Fig. 5. Frequency distribution of velocity perturbations of NOAA-5 fragments in the θ-x plane. The velocities are in the units of m/s.

Calculated variances of the velocity perturbations betray a most important finding of this study: the collision induced breakups have the least variances in Δv_θ and Δv_x and the low-intensity explosion induced breakups have the highest variances in Δv_θ and Δv_x whereas the high-intensity explosion induced breakups generally have variances intermediate between the two extreme classes. This is, of course, not inconsistent with the mechanisms of the breakups. A low-intensity explosion, like a pressure tank burst from deflagration in a Delta or Thor-Agena rocket produces fragments with a large velocity distribution whereas a high-intensity explosion like deliberate detonations of Cosmos series satellites produces fragments of more or less uniform velocity, i.e., with small velocity distribution (Benz, et al., 1987).

The above finding provides us with a scheme to separate the three types of satellite fragmentations: collision induced, low-intensity explosion induced and high-intensity explosion induced breakups. Figure 6 is a plot of the variances in Δv_θ and Δv_x with the variance of the plane change angle as the vertical axis. The plane change angles were taken from the study of Badhwar, et al. (1987). The positions of the satellites in the 3-space of Fig.6 clearly separate themselves out according to the following scheme:

Collision induced breakups .... low variances in Δv_θ, Δv_x and pca
High-intensity explosions ... Intermediate variances in the above
Low-intensity explosions ..... High variances in the above

The three satellites with the lowest variances were Cosmos 1375, Solwind P-78 and Cosmos 839, thus reaffirming the assertion that Cosmos 1375 and Cosmos 839 were target satellites in Soviet ASAT tests. Cosmos 1275, contrary to McKnight's analysis (Culp, 1986), places itself in the high-intensity explosion category as does Cosmos 61-63. Nimbus 4 and Transit 4A rockets have all the signs of belonging to the low-intensity explosion category.

It must be pointed out that the present study, like the Gabbard diagram

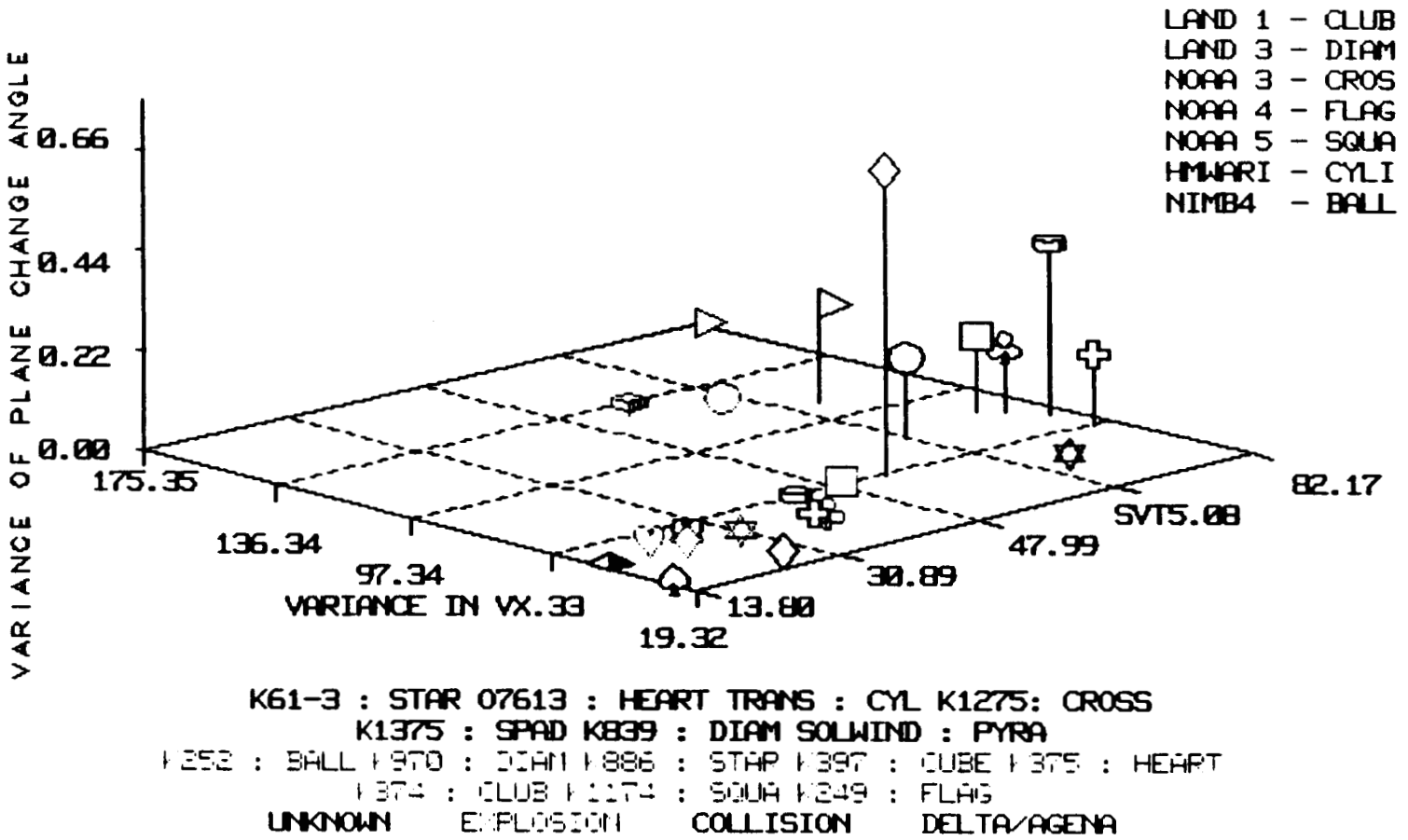


Fig. 6. Three-dimensional plot of variances of Δv_θ, Δv_x and pca.

studies of Culp and McKnight (Culp, 1986), suffers from the "ageing" of the fragment data. To obtain the best results, one needs the orbital elements of the parent just prior to the breakup as well as those of the fragments right after it. Unfortunately, this is seldom realized unless we can anticipate a breakup beforehand and make the necessary observations. In our examples, oftentimes the data were years old and several fragments had already decayed. The "youngest" data, on the other hand, often contained the fewest number of pieces. A judicious compromise always calls for a subjective decision of the user.

Lastly, we should seek alternative expressions for the velocity perturbations which do not pose such problems as encountered in this study. For example, in the case of nearly circular orbits, the radial component of the velocity perturbation may be approximated by the normal component, the expression for which is given by Ehricke (1962):

$$v_n = - \frac{a\ v\ \Delta e}{r \sin\theta} \quad . \tag{14}$$

Here, v is the tangential velocity of the parent at the point of fragmentation.

CONCLUSIONS

(1) An analysis of calculating velocity perturbations of fragments of a disintegrated satellite indicates that whereas the radial components frequently had abnormally high values, the components in the other two orthogonal directions had mostly acceptable values.

(2) The calculated variances in Δv_θ and Δv_x indicate that the collision induced breakups had the least variances, the high-intensity explosion induced breakups had higher variance while the low-intensity explosion induced breakups had the highest variances.

(3) A three-dimensional plot of the variances in Δv_θ, Δv_x and pca showed a clear separation in the three types of satellite breakups which can therefore be used to reclassify breakups of unknown causes.

(4) Cosmos 1375 and Cosmos 839 are thus reclassified as collision induced breakups; Cosmos 1275 and Cosmos 61-63 are reclassified into the high-intensity explosion category while Nimbus 4 and Transit 4A rockets are reclassified into the low-intensity explosion category.

(5) Other quantities from the satellite fragment data, such as angular momentum change, energy change, decay rates, etc. may also provide valuable information regarding the nature of satellite breakups and they could be undertaken as logical extension of this study.

REFERENCES

G. D. Badhwar, A. E. Potter, P. Anz-Meador and R. C. Reynolds, Preprint, 1987.

F. J. Benz, C. V. Bishop and M. B. Beck, Preprint, 1987.

T. D. Bess, NASA Tech. Note D-8108, Langley Research Center, 1975.

V. A. Chobotov, J. Spacecraft & Rockets, 20, 484, 1983.

R. D. Culp, Teledyne Brown Engg. Contract SC 7460, 1986.

K. A. Ehricke, Space Flight. II. Dynamics, Van Nostrand Co., Inc., Princeton, 1962, p.352.

W. Gellert, H. Kustner, M. Hellwich and H. Kastner, The VNR Encyclopedia of Mathematics, Van Nostrand Reinhold Co., New York, 1977, p.213.

N. L. Johnson, Communications with R. D. Culp, 1985.

N. L. Johnson, J. R. Gabbard, R. L. Kling, Jr. and T. W. Jones, Teledyne Brown Engg. Tech. Rept. No. CS86-USASDC-0001, 1986.

D. J. Kessler and B. G. Cour-Palais, J. Geophys. Res., 83, 2637, 1978.

R. Kling, Teledyne Brown Engg. Rept. CS86-LKD-001, 1986.

L. Meirovitch, Methods of Analytical Dynamics, McGraw-Hill Book Co., New York, 1970, p.456.

NORAD Space Surveillance Center (NSSC) Satellite Catalog, 1985.

R. C. Reynolds, Personal Communication, 1987.

L. Sehnal, COSPAR Meeting Presentation, 1984.

L. G. Taff, D. E. Beatty, A. J. Yakutis and P. M. S. Randall, COSPAR Meeting Presentation, 1984.

FINAL REPORT

Time-Domain Robotic Vision Application

NASA/ASEE Summer Faculty Fellowship Program-1987

Johnson Space Center

Prepared by: C. L. Tolliver, Ph.D.

Academic Rank: Professor

University & Department: Prairie View A&M University
Electrical Engineering Department
Prairie View, Texas 77446

NASA/JSC

Directorate: Engineering

Division: Tracking and Communications

Branch: Systems Analysis Office

JSC Colleague: Kumar Krishen, Ph.D.

Date: August 14, 1987

Contract Number: NGT 44-001-800

ABSTRACT

The quest for the highest resolution microwaves imaging and the principle of time-domain imaging has been the primary motivation for recent developments in time-domain techniques. With the present new technology fast time varying signals can now be measured and recorded both in magnitude and in phase. It has also enhanced our ability to extract relevant details concerning the scattering object.

In the past, the inference of object geometry or shape from scattered signals has received substantial attention in radar technology. Various inverse scattering theories were proposed to develop analytical solutions to this problem. Furthermore, the random inversion, frequency swept holography, and the synthetic radar imaging, all of which have two things in common: (a) the physical optic far-field approximation and (b) the utilization of the channels as an extra physical dimension, were also advanced significantly.

Despite the inherent vectorial nature of electromegnetic waves, these scalar treatments have brought forth some promising results in practice with notable examples in subsurface and structure sounding. The use of time-domain imaging for space robotic vision applications has been proposed. A multisensor approacah to vision has been shown to have several advantages over video-only approach.

FINAL REPORT

OBJECTIVES

The continued research efforts for this project should be as follows: (1) develop the application of time-domain imaging for robotic vision, in order to emphasize this we stress that it is the scattered field not the total power that is of interest; (2) research the hardware/technology needed for this application; (3) derive theoretical expressions for planar, cylindrical, spherical, and octobox configurations (the images will be synthesized on the basis of time-domain reflectance); (4) study the effects of pulse width, polarization, look angle, and carrier frequency and (5) compute magnitude and phase images. The limitations of the present analytical techniques will be identified. Also, experimental verification of the analytical results will be performed to develop hardware approaches.

INTRODUCTION

The detailed literature review conducted this summer at the Johnson Space Center, in the area of Time-domain microwave target imaging, led to a research topic of determining the size, shape, and electromagnetic properties of a scatter, given the incident and scattered eletromagnetic fields. The problem of interest here is the reconstruction of the shape of a convex scatter from knowledge of high frequency far field scattered from the object in response to a known incident field.

The quest for the highest resolution microwaves imaging and the principle of time-domain imaging has been the primary motivation for recent developments in time-domain techniques. In the last decade, modern techniques in sampling devices and advent of fast pulse generators has brought new technology to the practicing radar engineers in the measurements of picosecond at a much reduced cost. With the present new technology fast time varying signals can now be

measured and recorded both in magnitude and in phase.

Todays technology has enhanced our ability to extract relevant details concerning the scattering object, which conventional radar ranging lacks.

In the past, the inference of object geometry or shape from scattered signals has received substantial attention in radar technology largely due to its adademic significance and the understandable commercial and military values. In the last decade various inverse scattering theories were proposed to develop analytical solutions to this problem. Furthermore, the random inversion, frequency swept holography, and the synthetic radar imaging, all of which have two things in common:

(a) The physical optic far-field approximation.

(b) The utilization of the channels as an extra physical dimension, were also advanced significantly.

Despite the inheret vectorial nature of electromagnetic waves, these scalar treatments have brought forth some promising results in practice with notable examples in subsurface and structure sounding. Recently Dr. Kumar Krishen, of NASA/JSC, has proposed the use of time-domain imaging for space robotic vision applications. A multisensor approach to vision has been shown to have several advantages over video-only approach.

PROBLEM

(THEORETICAL APPROACH)

The problem of scattering from rough surfaces is inherently different in nature from that of scattering by other bodies. Usually in the rough-surface problem, an exact knowledge of the shape of the surface is neither available, nor is it of interest to the radar operator. Instead, only average properties of the surface shape enter into the problem. It is clearly understood that the last requirement rule out a boundry-value approach, since the exact boundry is not known. We are more interested in the relationship between the average scattered

field or radar cross section and the average surface properties. The radar cross section is normalized by the area A, and defined as, $\gamma(\theta_i, \theta_s, \varphi_s) = \langle\sigma(\theta_i, \theta_s, \varphi_s\rangle/A$, where the three angles θ_i, θ_s, and φ_s are shown for bistatic case in figure 1a below. The rough surface is assumed to consist of height variations about a mean plane, which is taken as the xy-plane. The incident wave lies in the xz-plane at a polar angle.

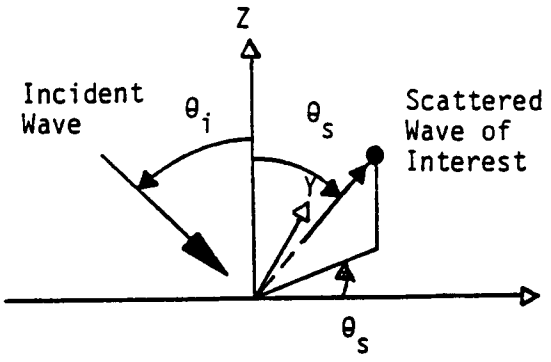


Figure 1a - BISTATIC SCATTERING GEOMETRY

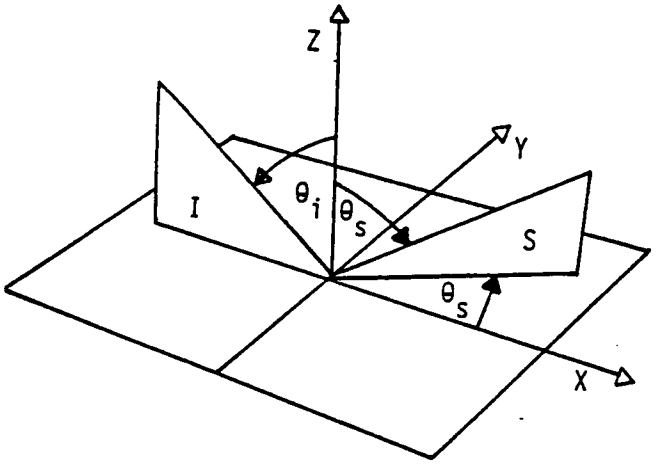


Figure 1b - THE "SCATTERING GEOMETRY"

Figure 1a shows the bistatic scattering for a planar rough surface whose mean

height coincides with the xy-plane and figure 1b shows the scattering geometry. I is the plane of incidence and S is the scattering plane.

The < > refer to an average. For the normaized average backscattering cross section are defined as

$$\gamma(\varepsilon_i) = \gamma(\varepsilon, \varepsilon, \pi)$$

Generally one has specific purposes or reasons for investigating scattering from a rough surface broadly these ultimate goals may be divided into three categories:

(1) The problem of direct scattering; one wishes to know the average properties of the scattered signal or cross section when the surface properties of the rough surface are known, and the scattering information is expressed in terms of the surface properties.

(2) The problem of inverse scattering; one wishes to obtain statistical information about the rough surfaces from a knowledge of the average properties of the scattered field. This problem is more difficult in that there appear to be many classes of rough surfaces producing the same average scattering cross section as a function of the bistatic scattering angle and wavelength.

The (scattered field) radar cross section, can be expressed as 4π times the power delivered per unit solid angle in the direction of the receiver divided by the power per unit area incident at the target. The factor 4π enters from the definition of solid angle. Assuming for the moment that the propagation path between the target and the receiving system is lossless, then this power ratio may be expressed as

$$4\pi r^2 (\Xi^s \cdot \Xi^{s*}) / (\Xi^i \cdot \Xi^{i*}) = 4\pi r^2 (\Xi^s \cdot \Xi^{s*}) / (\Xi^i \cdot \Xi^{i*})$$

where E^S and H^S are the scattered electric and magnetic fields, respectively and

E^i and H^i are the incident fields. The scattered field is defined to be the difference between the total field (with the targety present) and the incident field E^i (the total field that would exist if there were no target present). This is summarized as $E^S = E^T - E^i$.

Finding the cross section (σ) now becomes a problem in electromagnetic field theory. In order that the cross section be independent of r, it is desirable to let r in (Eg. 3) become arbitrarily large (scalar definition, not a tensor function).

$$\sigma = 4\pi \lim_{r\to\infty} r^2 (E^s \cdot E^{s*})/(E^i \cdot E^{i*}) = 4\pi \lim_{r\to\infty} r^2 (H^s \cdot H^{s*})/(H^i \cdot H^{i*})$$

To compute the scattered field one sometimes computes the current induced on the target and then treats the target current distribution in terms of an equivalent aperature distribution. Antennas are often compared to an isotropic antenna, that is, an antenna which radiates uniformly in all directions.

In the clutter problem, return from terrain is not wanted. However, in many cases the presentor clutter is unavoidably along with the desired signal, one can detect and analyze the desired signal significantly if more is known about the properties of the clutter or noise produced by terrain scattering. It is noted that surface information is generally known, and the properties of the scattered signal are sought related to (Eq. 3). At this point in time, a few more mathematical definitions may be of interest.

The source signal is related to the Fourier transform, F(w), by:

$$\text{(a) } F(w) = \int_{-x}^{x} f(t)e^{jwt}\,dt, \qquad \text{(b) } f(t) = \frac{1}{2\pi}\int_{-x}^{x} F(w)e^{-jwt}\,dt$$

Known by the Fourier transform pair as follows:

$$\text{(a) } F(w) = \int_{-\infty}^{\infty} f(t)e^{jwt}\,dt, \qquad \text{(b) } f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(w)e^{-jwt}\,dt$$

Capitals are used to denote transformed function unless otherwise noted or stated. The wave number k = w/c, where c is the speed of light in the medium. α and β are the direction cosines of the vector r projected onto the xy-plane (with $\alpha = \sin\xi\cos\xi$, $\beta = \sin\xi\sin\xi$).

It can be assumed that the field distribution over the antenna apertures and object plane is space - time separable; the field strength of each radiating element observed at some point r is $e(r,t) = e(r)f(t-\frac{r}{c})$ the field strength. Where $f(t-\frac{r}{c})$ is the sine wave traveling at the speed of light.

Consider the scatter in figure 2 to be placed in a far field, such that the assumption of a uniform impinging planewave over the object can be assured. The incident wave form i(t), on the object surface is obtained by the application of the far field Kirchhoff-Huygens principle, namely

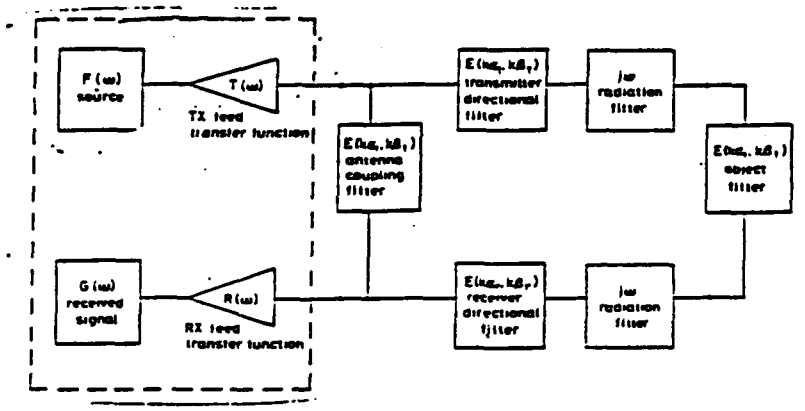


Figure 2.- SCHEMATIC DIAGRAM OF THE TARGET IMAGING SYSTEM

Let P be a point of observation and let R' be the distance from P to a point $\xi(x,y)$ on the surface S. The scattered field E at P is given by the Helmholtz integral as follows:

$$E(p) = \frac{1}{4\pi} \oint_s \left(E\frac{\partial \psi}{\partial n} - \psi\frac{\partial E}{\partial n}\right)ds$$

Let V be a volume bounded by a surface S and let A be any continuous vector function. Then, according to the divergency theorem,

$$\iiint_v \nabla \cdot A\,dv = \oint_s A \cdot n\,ds$$

Green's First Theorem:

$$\iiint_v \nabla E \cdot \nabla \psi\,dv = \oint E\frac{\partial \psi}{\partial n}ds - \iiint_v E\nabla^2\psi\,dv$$

Green's Second Theorem:

$$\iiint_v (E\nabla^2\psi - \psi\nabla^2 E)dv = \oint_s \left(E\frac{\partial \psi}{\partial n} - \psi\frac{\partial E}{\partial n}\right)ds$$

Note, the scattered field E(p) at the point of observation is given by the Helmholtz integral as follows:

$$E(p) = \frac{1}{4\pi}\oint_s \left(E\frac{\partial}{\partial n}\frac{e^{jkr}}{r} - e^{jkr}\frac{\partial E}{\partial n}\right)ds$$

(To get a better mathematical understanding of the application on how the scattered wave form is obtained, please refer to the author on the derivation of the scattered wave of interest in frequency domain).

$$I(w) = \frac{\cos\theta}{4\pi r_0} jwF(w)\int_x^x \int e(x,y)\exp\{-jk(r_0 - \alpha_x - \beta y)\}\,dxdy$$

$$I(w) = \frac{\cos\theta}{4\pi r_0} jwF(w)T(w)E_t(k\alpha_t, k\beta_t)\exp\{-jkr_0\}$$

Where $E_t(k\alpha_t, k\beta_t)$ is the transmitting-antenna anisortopic directional

filter and r is the distance between the transmitter aperture and an object plane taken as reference.

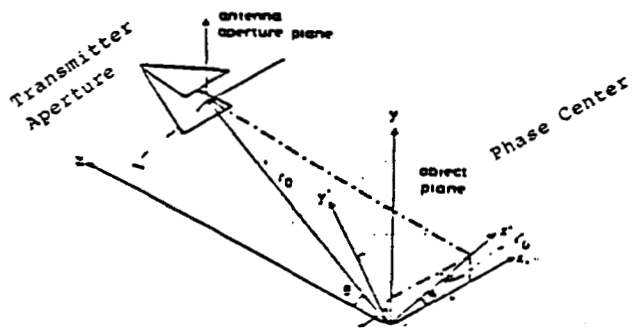
By stationary - phase evaluation and a second application of the Kirchhoff-Huygens principle, it can be shown, with little algebra that the received waveform, g(t), is

$$g(t) = \text{const}\int_{-\infty}^{\infty} F(w)T(w)R(w)[E_c(k\alpha_c, k\beta_c) + (jw)^2 E_1(k\alpha_1, k\beta_1)E_2(k\alpha_2, k\beta_2)E_0(k\alpha_0, k\beta_0)e^{-jkr}\,dw$$

For the monostatic case commonly encountered (see figure 3 below), the transmitter and receiver coincide. In particular, if the phase center of the object is on boresight which suggests that the target is closely tracted, eqr.16 can be considerably simplified to yield:

$$g(t) = \text{const}\int_{-\infty}^{\infty} F(w)T(w)R(w)[E_a(k\alpha_c, k\beta_c) + (jw)^2 E_1(0,0)E_2(0,0)E_0(k\alpha_0, k\beta_0)e^{-jwt}\,dw$$

The delineation of the object response is exercised as follows: First, the time response in the absence of any scattering objects is recorded.



(a) $X = \dfrac{\partial x' + \beta y'}{\sqrt{\partial^2 + \beta^2}}$ (b) $y = \dfrac{\beta x' - \partial y'}{\sqrt{\partial^2 + \beta^2}}$

(c) $\text{Cos}\,\phi = \dfrac{\partial}{\sqrt{\partial^2 + \beta^2}}$ (d) $\text{Sin}\,\phi = \dfrac{\beta}{\sqrt{\beta^2 + \partial^2}}$

Figure 3 - IMAGING GEOMETRY

THE ROTATION TRANSFORMATION

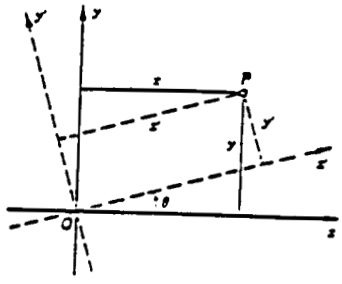


FIGURE 3a.

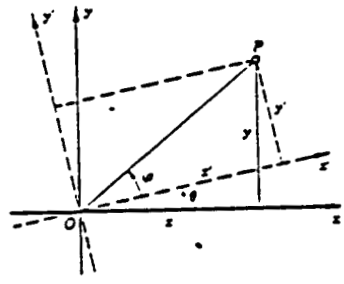


FIGURE 3b.

(f) $x = x' \cos\theta - y' \sin\theta,$
$y = x' \sin\theta + y' \cos\theta,$

(g) $x' = x \cos\theta + y \sin\theta,$
$y' = -x \sin\theta + y \cos\theta.$

(h) $\frac{x}{|OP|} = \cos(\varphi + \theta),$
$x = |OP| \cos(\varphi + \theta)$
$= |OP| (\cos\varphi \cos\theta - \sin\varphi \sin\theta).$

(i) $\sin\varphi = \frac{y'}{|OP|}$ and $\cos\varphi = \frac{x'}{|OP|}$

This signal is subsequently substracted from other records with the scatter present, thus removing the antennas coupling filter. Given a sufficiently large free space, multiple path scattering between the object and surrounding obstacles can be gated out in time. By definition

$$\Xi_0(k\alpha_0, k\beta_0) = \int_{-\infty}^{\infty}\int e(x,y)\exp[+jk(\alpha_0 x + \beta_0 y)]\,dx\,dy$$

Having removed the coupling term, the backscattering waveform g(t), gives the

system kernel, K(w),

$$K(w) = \cos t + (jw)^2 F(w)T(w)R(w)\Xi_r(0,0)\Xi_r(0,0)e^{-2jkr_0}$$

with the use of mathematical techniques, the backscattering waveform was simplified, to give the system kernel, K(w), further giving an opportunity to obtain the electric field strength in frequency domain

and by the Projection theorem , the response obtained is equivalent to a slice in the two dimension fourier plane (Figure 4).

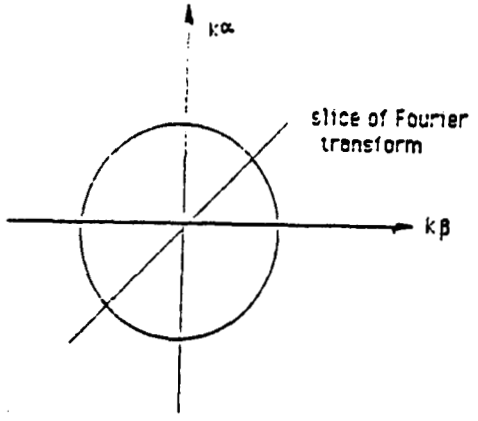


Figure 4. - REPRESENTATION OF THE ONE-DIMENSIONAL TARGET FREQUENCY RESPONSE IN THE TWO-DIMENSIONAL FOURIER PLANE AT A PARTICULAR VIEWING DIRECTION

It is further noted that the help of figures 3a, 3b assisted in a change in spatial coordinates in Figure 3 from one dimension coordinate system to the two dimension coordinate system $\Xi(k\alpha, k\beta)$.

Followed by an inverse fourier transform (Eq. 11) yields the object impulse response $\int_{-\infty}^{\infty} \Xi(k,o)e^{-jw(t-2x'\sin\frac{\xi}{c})}dw = e(t - 2x'\sin\frac{\xi}{c})$

which can be considered as a series of scans across the object surface where each is an integral of the object field which is orthogonal to the viewing direction. The spatial resolution is thus achieved by the utilization of the time channel and the minimal resoluable dimension Δx is related to the sampling integral T, namely $\Delta x = \frac{c\pi}{2\beta \sin\theta}$

The two-dimensional object field distribution is recovered, namely the reconstruction field strength, e(x,y)

$$e(x,y) = \frac{1}{4\pi^2} \int_0^{\pi} \int_{-\infty}^{\infty} \Xi(k\alpha, k\beta) A(k\alpha, k\beta) e^{j k(\alpha x, \beta y)} k d\alpha d\beta$$

a smoothing filter $A(k\alpha, k\beta)$ is introduced to compensate due to the factor k in the field strength equation $e_0(x,y)$. It is not necessary to have A, for noise free data. For practical numerical evaluation, the finite sampling rate will impose an upper limit, R, to the wave number k. To eliminate abrupt cutoff which would cause ringing in the resultant image, it is necessary to introduce

$$A(k\alpha, k\beta) = \sin\left(\frac{k\pi}{2R}\right) / \left(\frac{k\pi}{2R}\right)$$

which yields a smoothed image

$$e(x,y) = \frac{1}{4\pi^2} \int_0^{\pi} d\beta \int_{-R}^{R} \frac{k}{2} \Xi(k\alpha, k\beta) \sin\left(\frac{k\pi}{2R}\right) e^{j k(\alpha x, \beta y)} dk$$

and is the basis of our reconstruction field strength.

RESULTS

Theoretical calculations of the single and train pulse of the schematic diagram (figure 2) of the target imaging sytem led me to analyze fourier transform of the input function. A program was written in Basic and the use of BOSS (Block Oriented System Simulator) enable me to plot the energy supplied to the target and give the time plot of a single and train of pulses. I was also able to verify the magnitude spectrum of a single pulse continuous descrete function and the magnitude spectrum of a line spectrum of pulse train.

APPLICATIONS

- Recently NASA/Johnson Space Center has proposed the use of time-domain imaging for Space Robotic vision applications.

- Many space objects are covered with heat shields, etc. (Dielectric Materials) and microwave can sense through this layer.

- A multisensor approach to this vision has been shown to have several advantages over video-only approach.

RECOMMENDATIONS

- It is recommended that the second phase of the research project, inititated at the NASA/Johnson space Center, in the area of Time-Domain Robotic Visions Applications, be continued to study the effects o9f pulse width, polarization, look angle, and phase images on robotic vision, and to conduct experiments to verify the theoretical results.

- The amount of time provided for the theoretical and experimental verification of the Time-Domain Microwaive Robotic Vision Application was not sufficient in order to perform the above experimental verification.

$$|F(w)|^2 = \frac{\pi^2}{\tau_0^2}\left[\frac{1}{\omega}+\sum_{n=1}^{\infty}\frac{\sin\frac{n\pi}{\tau_0}}{\frac{n\pi}{\tau_0}}\,\frac{\omega}{\omega^2-\left(\frac{n\pi}{\tau_0}\right)^2}\right]\left[\frac{1}{\omega}+\sum_{n=1}^{\infty}\frac{\sin\frac{n\pi}{\tau_0}}{\frac{n\pi}{\tau_0}}\,\frac{\omega}{\omega^2-\left(\frac{n\pi}{\tau_0}\right)^2}\right]$$

```
LIST
5 PRINT "The Energy Supplied To The Target is |F(w)|^2 = F(w)F(w)*"
6 PRINT
7 PRINT
8 PRINT "|F(w)|^2="
9 PRINT
10 FOR N = 1 TO 4
15 PRINT  "N","w","F(w)F(w)*"," B",
16 PRINT
20 FOR W = 5 TO 10
30 B = (2/( N * 3.14 ))* (SIN (N * .0314)) * (W/( W*W - N*2*3.14)/EXP(10-10))
35 FW = ( 1 /( 2 * 3.14)* (.01))+ B * (1/(2*3.14) *(.01))-B
60 PRINT  N,W,FW,B
70 NEXT W
80 NEXT N
O
```

```
RUN
The Energy Supplied To The Target is |F(w)|^2 = F(w)F(w)*

|F(w)|^2=

N             w             F(w)F(w)*                B
 1             5             6.941864E-03             5.341003E-03
 1             6             5.635806E-03             4.037021E-03
 1             7             4.874189E-03             3.276615E-03
 1             8             4.368317E-03             2.771547E-03
 1             9             4.004789E-03             2.408598E-03
 1             10            3.729419E-03             2.133665E-03
N             w             F(w)F(w)*                B
 2             5             9.63845E-03              8.033302E-03
 2             6             6.716592E-03             5.116089E-03
 2             7             5.437877E-03             3.839407E-03
 2             8             4.705682E-03             3.108376E-03
 2             9             4.224854E-03             2.628312E-03
 2             10            3.881775E-03             2.285779E-03
N             w             F(w)F(w)*                B
 3             5             1.782794E-02             1.620977E-02
 3             6             8.586145E-03             6.982669E-03
 3             7             6.234785E-03             4.635048E-03
 3             8             5.135711E-03             3.537721E-03
 3             9             4.488433E-03             2.891472E-03
 3             10            .0040569                 2.460625E-03
N             w             F(w)F(w)*                B
 4             5             -.8308692                -.8311381
 4             6             1.261031E-02             1.100044E-02
 4             7             7.448911E-03             5.847244E-03
 4             8             5.703307E-03             4.104415E-03
 4             9             4.810195E-03             3.212726E-03
 4             10            4.260522E-03             2.663923E-03
O
```

Table I - The Energy Supplied To the Target

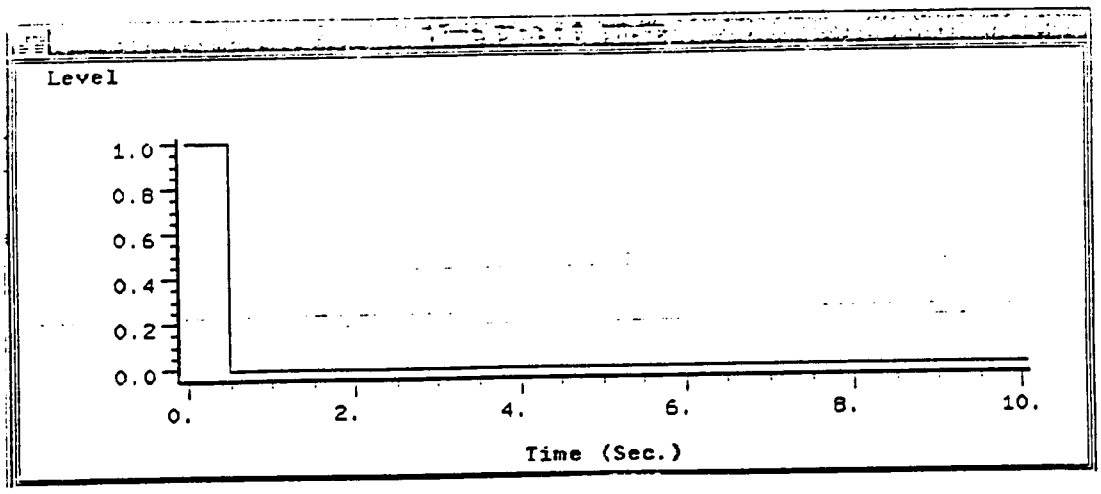


Figure 5.- Time Plot of a Single Pulse

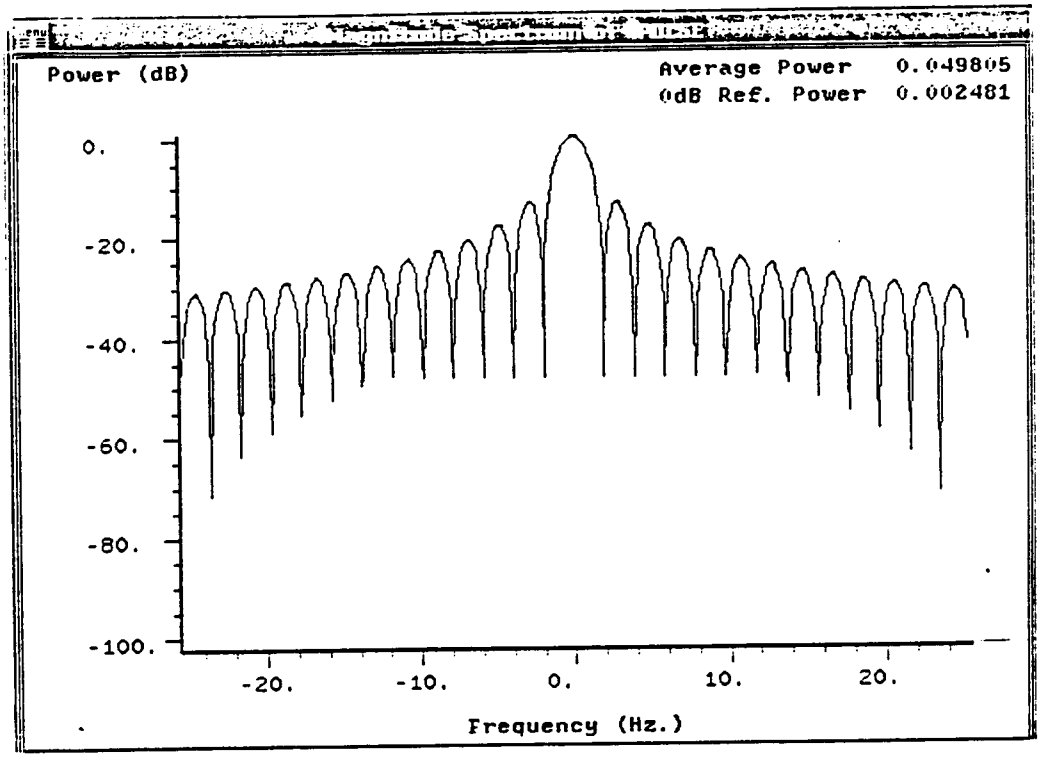


Figure 6.- Magnitude Spectrum of a Single Pulse

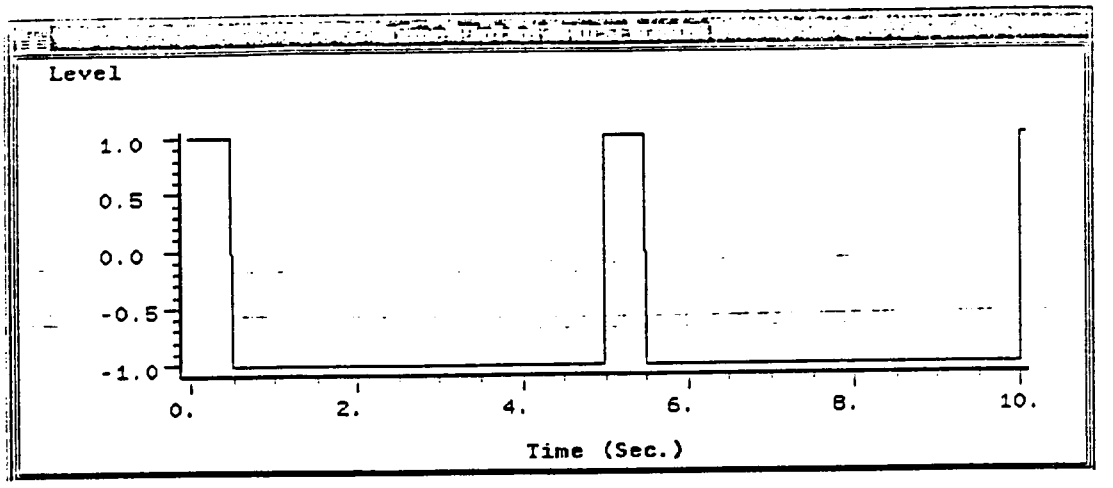


Figure 7.- Time Plot of A Pulse Train

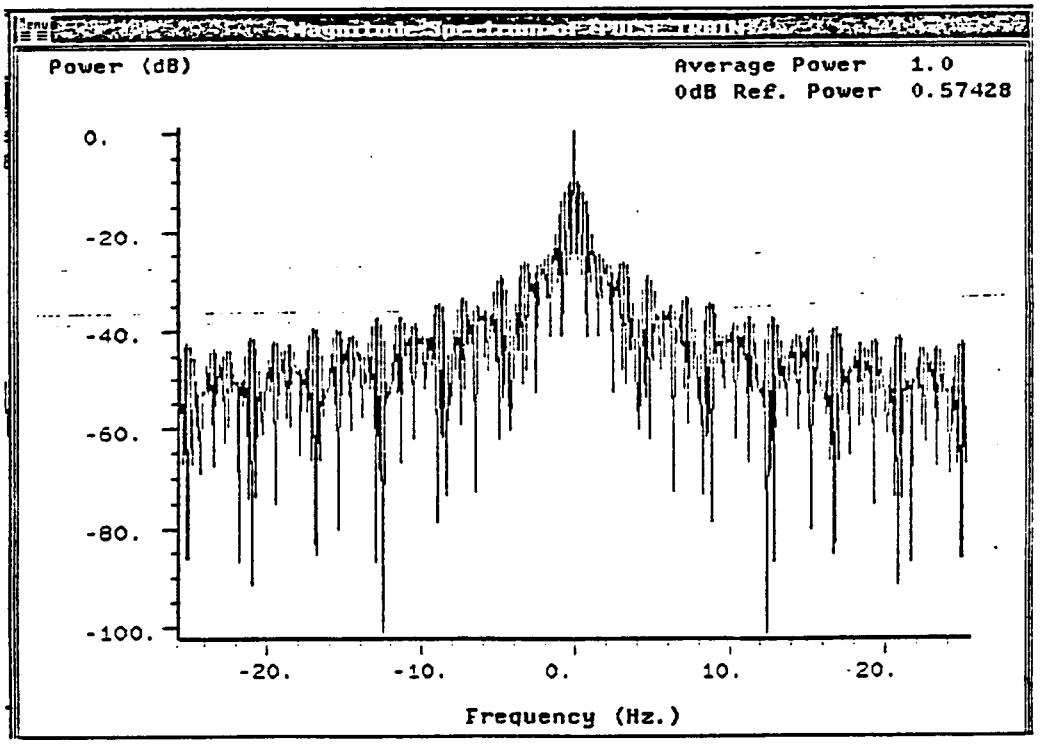


Figure 8.- Magnitude Spectrum of A Pulse Train

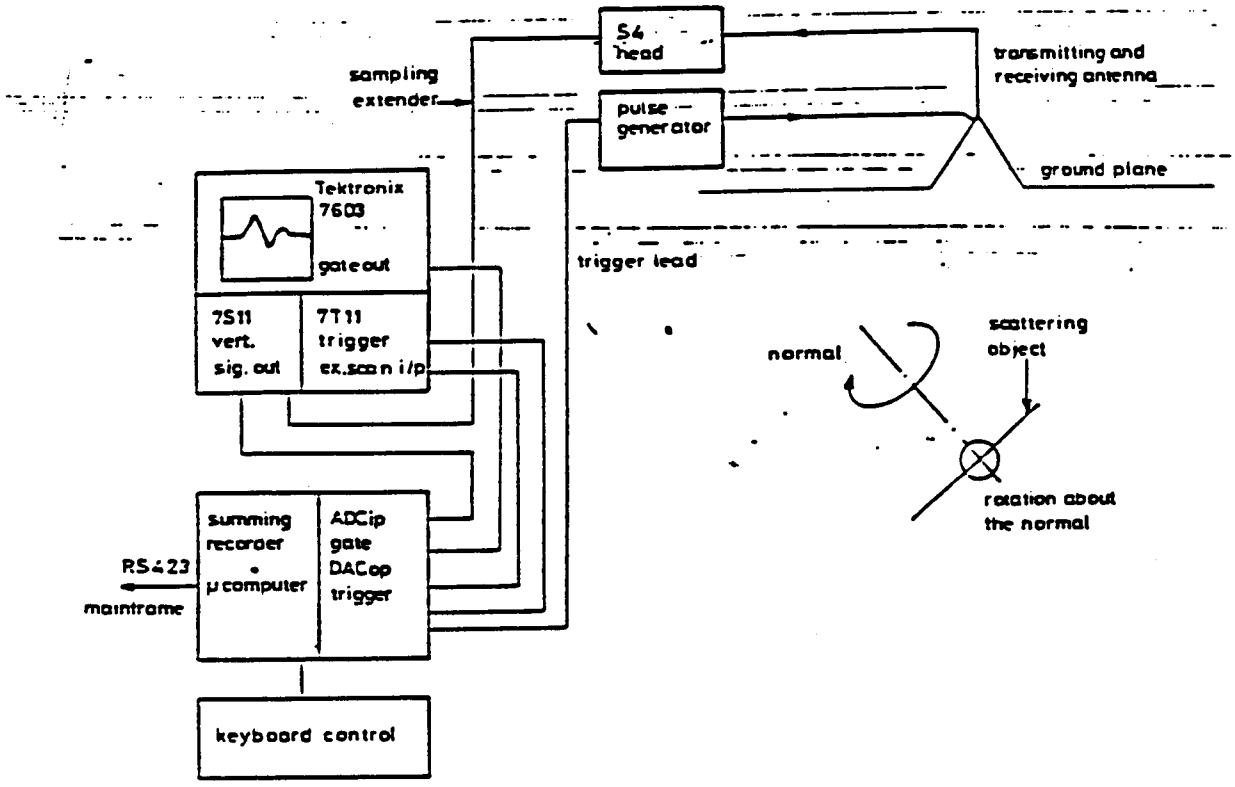


Figure 9.- Experimental Arrangement of the Target Scattering Measurement

The above is a circuit diagram which will be used to verify the results of elliptical and bowtie geometry of targets (Figures 10, 11, 12, and 13)

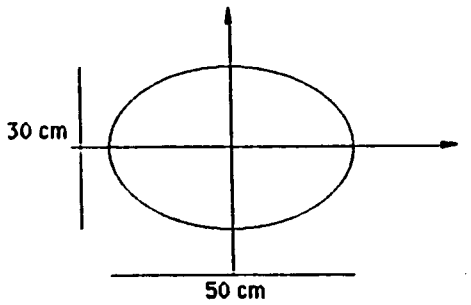


Figure 10- Geometry of the elliptical target.

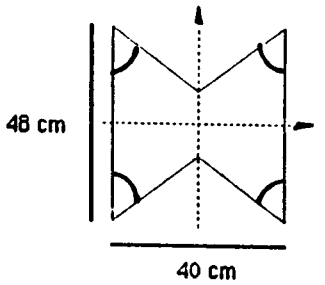


Figure 12- Geometry of the Targets - Bow-Tie

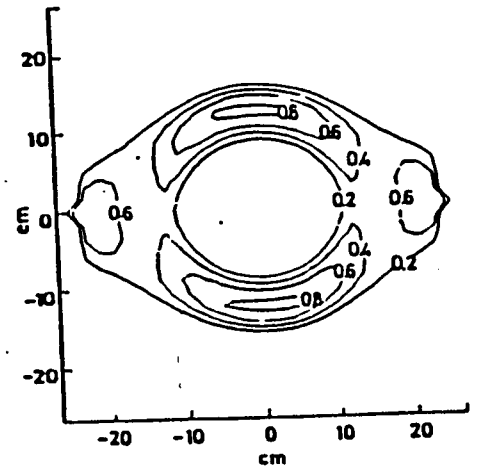


Figure 11- Reconstructed Images -Ellipse

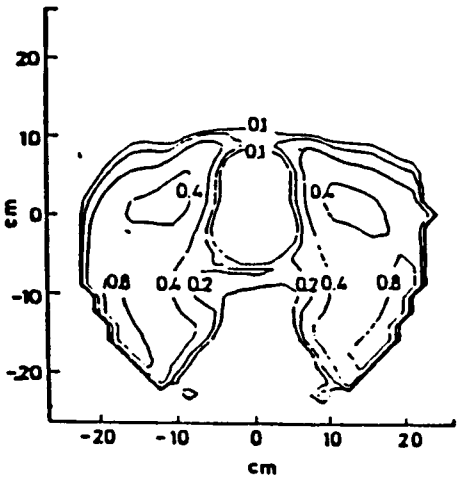


Figure 13- Reconstructed Images - Bow-Tie

ACKNOWLEDGEMENT

This research was conducted under the direction of Dr. Kumar Krishen and the Tracking and Communications Division (TCD). I also enjoyed discussions and interaction with several TCD personnel including James E. Ratliff, Ricardo J. Villarreal, Budy J. Walker and Ned J. Robinson. The encouragement of Mr. R. S. Sawyer, Chief, TCD is greatly acknowledged. Special acknowledgement to Ms. Michelle K. Morris for her typing support.

BIBLIOGRAPHY

Bachman, C.G.: Radar targets(Gover Pub. Co. Ltd., 1982), pp. 110-113.

Beckman, P. and Spizzichino, A.: The Scattering of Electromagnetic Waves from Rough Surfaces(The Macmillan Co., 1963), pp. 10-28, 178-181.

Bennett, C.L., and Ross, G.F.: Time domain electromagnetics and its applications, ibid., 1978, 66, pp. 299-318.

Boerner, W.M., HO, C.M., and Foo, B.Y.: Use of Radon's projection theory in electromagnetic inverse scattering, IEEE Trans., 1981, AP-29, pp, 336-341.

Chan, C.K., and Farhat, N.H.: Frequency swept tomographic imaging of three dimensional perfectly conducting objects, ibid, 1981, AP-29, pp. 312-319.

Censor, Y.: Finite series-expansion reconstruction methods, ibid, 1983, 71, pp. 409-419.

Evans, S., and Kong, F.N.,: Gain and effective area of impulse antenna, Third International Conference on Antenna and Propagation, ICAP 83, Norwich, England, April 1983, pp. 421-424.

Fialkovskiy, A.T.: Diffraction of planar electromagnetic waves by a slot and a strip, Radio Eng. Electron., 1966, 11, pp. 150-157.

Kennaugh, E.M. and Moffatt, D.C.: Transient and impulse approximations, Proc. IEEE, 1965, 53, pp. 893-901.

Kong, F.N.: Ph.D. Dissertation, Cambridge University, August 1983, Chap. 8.

Lewitt, R.M.: Reconstruction algorithms: transform methods, Proc. IEEE, 1983, 71, pp. 390-408.

Reader, H.C., Evans, S., and Yeaung, W.K.: Illumination of a rectangular slot radiator over a 3 Octave bandwidth, Fourth International Conference on Antennas and Propagat., ICAP 85, 1985, pp. 223-226.

Ruck, G.T., Barrick, D.E., Stuart, W.D., and Krichbaum, C.K.: Radar Cross Section Handbook(Plenum Press, 1970), pp. 671-689.

Schubert, K.A., Young, J.D., and Moffatt, D.L.: Synthetic Radar imaging, ibid, 1977, AP-25, pp. 477-483.

Tikhonov, A.N. and Arsenic, V.Y.: Solutions of ill-posed problem (Winton-Wiley, New York, 1977).

Yeung, W.K. and Evans, M.A.: Time-domain microwave target imaging, Proc. IEEE, 1985, 132, pp. 345-350.

A STUDY OF THE MINORITY COLLEGE PROGRAMS AT THE NASA JOHNSON SPACE CENTER

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1987
Johnson Space Center

Prepared by:	Mfanya Donald Tryman, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Prairie View A&M University Benjamin Banneker Honors College Prairie View, Texas 77446
NASA/JSC	
Program Office:	Equal Opportunity Programs Office
JSC Colleague:	Joseph D. Atkinson, Jr., DPA
Date:	August 14, 1987
Contract Number:	NGT 44 001 800

ABSTRACT

This project examined research programs in science and engineering at predominantly Black and White universities which assist in training and furthering the capabilities of minorities in the field. The Minority Graduate Researcher's Program and the Historically Black College and University Program were the focus of this research. The obejectives included investigating the organizational structure and processes of the programs, how they are run, how they differ, defining particular administrative tasks for these programs, the collection of data related to these programs, and recommending ways in which these programs can be improved for greater efficiency and effectiveness through the Equal Opportunity Programs Office.

INTRODUCTION

The Johnson Space Center (JSC) in Houston, Texas has made significant progress in diversifying employee personnel based upon race, national origin, religion, and gender. (1) This has been particularly true over the last fifteen years, in which the percent of minorities has more than doubled. (2) Nevertheless, in recent years the number of Blacks in particular graduating with degrees in science and engineering has experienced a significant decline. (3) The Johnson Space Center, along with other Centers within the National Aeronautics and Space Administration (NASA), attempted to remedy this situation with innovative programs. The two primary approaches utilized to achieve this end are the Historically Black Colleges and Universities (HBCU) program, and Minority Graduate Researchers (MGR) programs. These programs have been developed recently with the expressed purpose of actively engaging more minorities in NASA research and development projects.

THE HBCU PROGRAM

The HBCU program targets historically Black schools for research grants from NASA Centers. These grants provide monies in the areas of science and engineering for faculty at HBCU schools to conduct research with the assistance of their graduate students. It is understood that the awarding of grants to HBCU faculty carries with it the responsibility to use and pay students for their services as research assistants. NASA feels that this will provide an invaluable experience for both present and future Black scientists and engineers who may, in turn, be recruited by the space agency or otherwise contribute to the field. President Reagan noted the significance of HBCU's in education and training Blacks on September 20, 1986, in a presidential proclamation and designating September 15-21, 1986, as "National Historically Black Colleges Week." (4) Prior to Reagan's proclamation, on September 15, 1981, the president issued Executive Order 12320 "which mandates a significant increase in the participation of HBCU's in Federally sponsored programs." (5) The Order requires federal agencies to submit annual plans with tangible objectives regarding proposed agency actions to assist HBCU's.

As Table I reveals, for FY86 (as of 7/31/86), NASA had awarded 5.7 million dollars to HBCU's for scientific and engineering research related to NASA projects. This was inclusive of some 16 colleges and universities located throughout the South.

TABLE I: FISCAL YEAR 1986 NASA HBCU RESEARCH AWARDS

School	Amount
Alabama A&M	$377,000
Atlanta	280,000
Bowie State	127,000
Fisk	243,000
Florida A&M	375,000
Hampton	696,000
Howard	1,132,000
Morgan State	337,000
Norfolk State	33,000
N.C. A&T State	747,000
Prairie View A&M	208,000
Southern, Baton Rouge	272,000
Tennessee State	160,000
Tuskegee	356,000
U. Maryland-Eastern Shore	85,000
Virginia State	200,000

Source: Office of the Chief Scientist, NASA Headquarters, October 3, 1986.

Awards totaled as much as $1,132,000 for Howard University and as little as $33,000 for Norfolk State. While these schools represent only about 15% of the 110 HBCU's, it is a positive initiative. This initiative in FY85 allowed 150 graduate students and 235 undergraduate students at HBCU's to participate in NASA-related programs.(6) For FY1986, eight million dollars were earmarked for HBCU's and available through NASA Centers. As Chief Scientist Frank B. McDonald notes:

> We now plan to encourage principal investigators who have NASA research grants and a need for further student involvement to seek out talented underrepresented minority students and involve them in their NASA projects. The underrepresented minorities who will be the special focus of this effort will be Blacks, Hispanics, American Indians and Pacific Islanders.(7)

THE MGR PROGRAM

The MGR program focuses, like the HBCU program, on underrepresented minority students defined as Blacks, Hispanics, American Indians and Pacific Islanders. Unlike the HBCU initiative, the program is not limited to Black schools. The program targets minority students who may be at predominantly Black or predominantly white colleges and universities. Students in this program must be enrolled full-time working on an advanced degree in science or engineering but may be considered as a candidate prior to receiving their baccalaureate degree.(8) Successful candidates in the MGR program may receive awards up to $18,000 a year renewable for a total of three years culminating in an advanced degree.(9) These awards may be given separately to students apart from NASA funding for a research grant in cases in which the student cannot be accommodated in the original grant.

The MGR program evolved, in part, as a follow-up to the HBCU initiative and President Reagan's Executive Order. The groundwork for the Minority Graduate Researchers program was laid in October 1984, when the NASA Advisory Committee on Minority Graduate Researchers (NAC/MGR):

> recommended that NASA initiate and expand its research relationship with those universities, add underrepresented minorities to its existing principal investigators grants, and get more underrepresented minorities involved in other relevant NASA activities. The E.O. Council concurred in these recommendations.(10)

The MGR program has, among other objectives, the "development of minority talent" as "an effective way to utilize the potential of this nation's citizenry; and it helps increase the resource pool of research skills that will be needed to meet aerospace and other technological objectives of the future."(11) Because the MGR program is more diversified and flexible than the HBCU, it allows NASA to reach out to minority students who may be at a non-HBCU school that has a NASA research grant. For FY87, a two million dollar budget was earmarked for the MGR program through NASA's Offices of Equal Opportunity Programs and the Education Division at Headquarters.(12)

With this cursory overview of NASA's minority research programs at the college level, the remainder of this study will examine the JSC organization chart, the organizational structure of the HBCU and MGR within JSC, suggest the particular tasks/roles that the E.O. unit should perform in the college program, investigate how minority proposals from HBCU's are processed at JSC, evaluate the programs that the EOPO at the JSC currently holds and the status of research grants at the respective schools, and recommend ways and means in which the HBCU and MGR programs can be improved upon.

METHODOLOGY

This study was conducted from mid-May 1987 through early August 1987. In addition to the traditional literature search and review, the methodological approach for this study involved the collection of various and sundry data related to the HBCU and MGR program at the Johnson Space Center. This involved, firstly, constructing an instrument composed of relevant categories of data needed that were related to the respective programs. Secondly, phone numbers of the various actors involved in the HBCU and MGR programs had to be obtained. And thirdly, data were actually collected utilizing phone surveys to fill the relevant categories. This latter task proved to be somewhat difficult.

Some of the principal investigators, technical monitors, and buyers were either not available, out of town, or simply did not return phone calls that were placed to collect the information. Because these data were already fragmented between different individuals, this only exacerbated the task. As a result, a significant part of the months of June and July were spent in collecting these data. A formal assessment of the programs was made based upon the data that were collected. Organizational charts, tables, and figures were taken from primary (NASA JSC) and secondary sources (periodicals). The summary, conclusion, and recommendations capture the most important components of the study for future direction of these programs at the NASA JSC.

ORGANIZATIONAL STRUCTURE

The organizational structure at the Johnson Space Center is composed of a director, deputy director and executive assistant at the top of the administrative hierarchy. At the next level, there are six units that report directly to the center director which are inclusive of personnel, plans, safety, reliability and quality assurance, chief counsel, public affairs and the equal opportunity program. As revealed in Figure 1, there are a number of other divisions located within the hierarchy of the center. It is the Equal Opportunity Programs unit that is responsible for administering the HBCU and MGR programs. The Equal Opportunity Programs unit functions as a staff rather than line unit in the organizational structure.

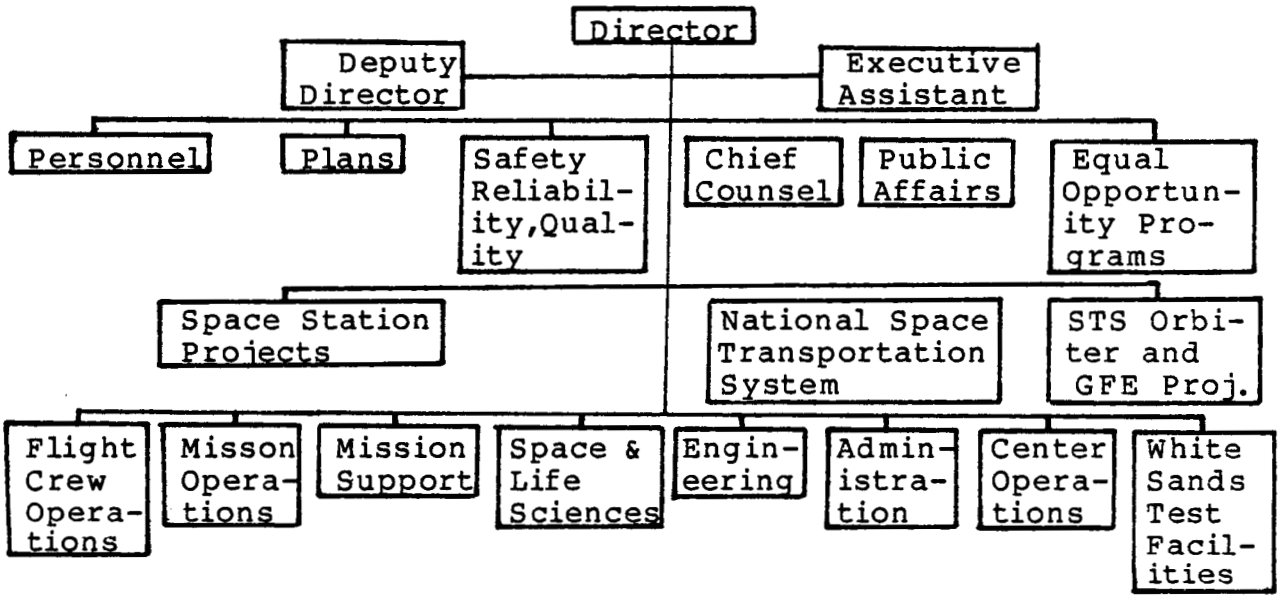


FIGURE I: NASA JSC ORGANIZATIONAL STRUCTURE
Source: Johnson Space Center, 1987

The organizational chart in Figure I constitutes personnel in the respective units that make up the policymaking body of the JSC as well. Because the Chief of the Equal Opportunity Programs reports to the center director and represents a staff division, the division is in a position to carry out policies coming straight from the top of the administrative hierarchy. Figure II shows one way in which the hierarchical structure of the HBCU and MGR programs may be conceptualized at the JSC. The EO unit,

while reporting directly to the center director, is responsible for identifying the various needs of the JSC divisions and providing an isomorphic match with HBCU's and other schools participating in the MGR program. In Figure II, the EO office becomes the focal point of attention because of its vital role in the overall programmatic goals.

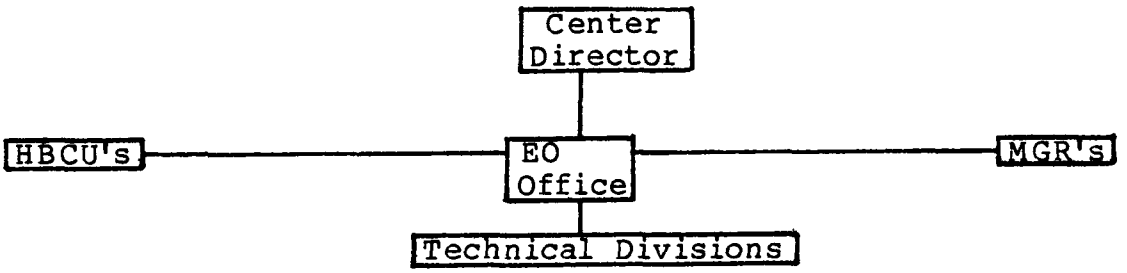


FIGURE II: HIERARCHICAL STRUCTURE OF COLLEGE PROGRAMS
Source: Johnson Space Center, 1987

While Figure II may be appropriate from a conceptual standpoint, it does not explain the specific functions or tasks of the EO Office in the overall process. This appears to be one of the major weaknesses of the college programs at the current time. There is a need to delineate the particulars with regard to the various roles that the EO Office is to perform. While it is the duty of the EO Office to coordinate NASA funding with proposals from institutions of higher learning, this task is too generic. Consequently, there is a need to develop other tasks and strategies for the EO Office in order to make the college programs successful.

TASKS TO BE PERFORMED

It is recommended that the following tasks be performed by the EO Office with regard to the college programs:

1.) develop a game plan and strategy for maximizing the utilization of JSC personnel who can respond affirmatively in implementing the HBCU and MGR programs,

2.) explore the possibilities by JSC of including more institutions of higher learning in programmatic goals of the HBCU and MGR objectives,

3.) systematically identify more HBCU's that have programs, departments or divisions in science and engineering and increase communications and public relations with these schools by intensifying letter writing, campus visits, etc.,

4.) a needs assessment of the various divisions at the JSC should be conducted to determine in a systematic manner those projects that are potentially compatible with the scientific and engineering missions of HBCU's, and encouraging these institutions to write related proposals to receive grants and contracts,

5.) institutions of higher learning with existing NASA research grants and contracts should be contacted for continuation and/or expansion where such institutions have a history of success and project completion,

6.) construct recruitment and visitation committees to make site visits to HBCU's,

7.) contact predominantly white institutions of higher learning that have NASA grants and contracts to determine if they have an interest in identifying minority research assistants as part of the MGR program, and

8.) meet with the technical monitors at JSC on a bi-monthly basis who are assigned to the various grants and contracts in the HBCU and the MGR program.

It is further recommended that the above tasks be carried out and conducted on a regular basis. In short, these tasks should be institutionalized in order to optimize and insure a higher degree of success. At present, some of the above tasks are either non-existent or not carried out on a regular basis. The institutionalization of these tasks appear to be an imminent necessity.

THE UNSOLICITED PROPOSAL PROCESS

The flow chart found in Figure III below suggests the manner in which unsolicited proposals from HBCU's are processed at the JSC which ultimately lead to the acceptance

and approval of a research grant or the rejection and disapproval of the same. Since most of the emphasis is upon (though not exclusively) the HBCU program, only a flow chart addressing this program is offered at this time.

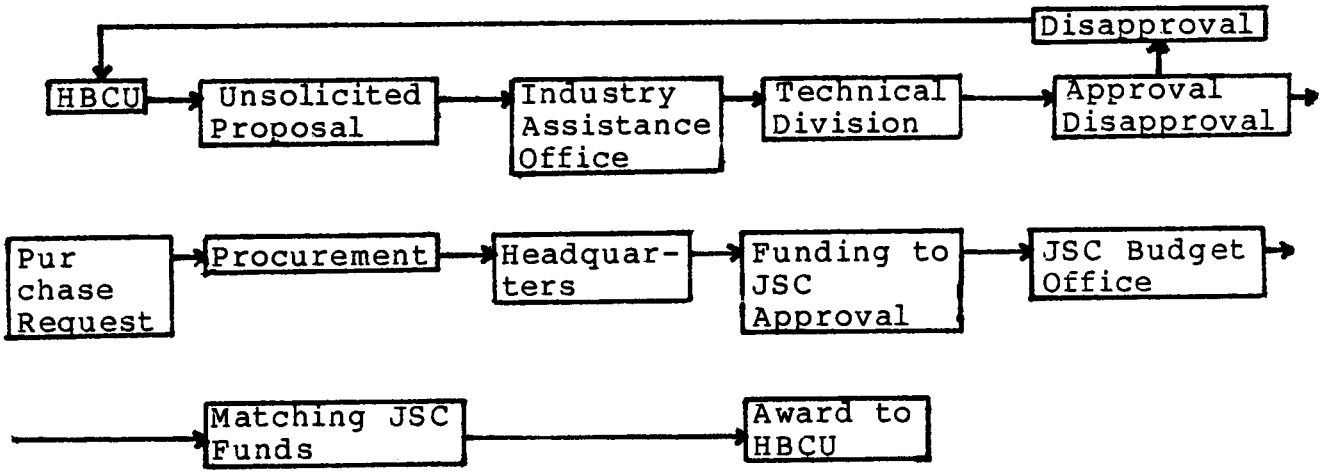


FIGURE III: THE UNSOLICITED PROPOSAL PROCESS
Source: Procurement Office, Johnson Space Center, 1987

In Figure III, HBCU's are encouraged to submit research proposals for funding from the JSC. Because these proposals are unsolicited as noted in the flow chart, Black institutions of higher learning are allowed to apply as the "sole source" regarding grant funding. Unsolicited proposals are sent to the Industry Assistance Office at JSC through the E.O. Office. The Industry Assistance Office, in turn, sends the proposal to the appropriate technical division related to the specifics of the proposal. It is the responsibility of the technical division to approve or disapprove the proposed study. If the division disapproves the proposed study, the proposal is returned to the principal investigator at the HBCU in question. On the other hand, if the proposal is approved, a purchase request is prepared and sent to procurement. In turn, the proposal is sent to headquarters for approval which provides one-half of the funding for the project to the JSC. The proposal, along with one-half of the funds, is sent to the JSC Budget Office which provides the other one-half or matching funds. Once the proposal has completed this process the research grant award is made to the HBCU. It is important that the EO Office closely monitor the process as proposals are submitted and travel through the various steps outlined herein.

EVALUATION OF EXISTING GRANTS/CONTRACTS

This study has attempted to systematically examine a total of four schools that have research grants or contractsfrom the NASA-JSC as evidenced in Appendixes A and B. Of these four schools, three are HBCU's and one is a predominantly white institution of higher learning. The categorical breakdown of schools in these appendixes is an attempt to elicit a complete picture of the nature of the programs in question. It reflects data and information most pertinent in running the HBCU and MGR programs efficiently and effectively. Several points are worthy of mention with regard to these data and findings.

One, while there are only three HBCU's with NASA research grants at the Johnson Space Center, Prairie View has been successful in having four separate research projects funded, although one of these is directly from NASA Headquarters. When one considers the fact that there are some 110 HBCU's in the country, nevertheless, this is a rather modest number of schools as well as contracts. The diversity and expansion of these programs to other HBCU's would be a welcome addition.

Two, it is important to expand the MGR program to other predominantly white institutions of higher learning. Although the number of Blacks enrolling in science and engineering programs are declining, more are attending these colleges and universities while less are attending HBCU's. In one context, this presents a dilemma because HBCU's in almost every case are underfunded while predominantly white schools remain solvent in most cases. This is true of public as well as private HBCU's, of which the latter is revealed in Table II.

TABLE II: WEALTHIEST PRIVATE HBCU'S

College/University	Endowment
Spelman College	$33,381,800
Tuskegee University	$28,452,400
Dillard University	$14,332,812
Morehouse College	$12,674,146
Benedict College	$12,588,996
Atlanta University	$11,254,011
Jarvis Christian College	$11,103,871
St. Augustine's College	$9,088,560
Stillman College	$9,005,209
Rust College	$7,151,398
Total:	$149,032,945

Source: United Negro College Fund, 1985-86 Academic Year

In contrast to these schools listed above, the total of over 149 million dollars is merely a drop in the bucket compared to those endowments of predominantly white schools. Harvard, for instance, alone has an endowment of more than $3.45 billion.(13)

One of the primary goals of NASA research grants to HBCU's is to prepare future minority scientists and engineers who could potentially be hired by NASA once they have received the terminal degree or contribute to their fields in other ways. A cursory analysis of the data in the appendixes leads one to the conclusion that this goal is being fulfilled. In the case of each research grant and/or contract, anywhere from one to 30 students have participated in one or more phases of the project. Some students have served as paid research assistants while others have utilized their work as part of their own research project, i.e., a masters thesis. The utilization of, in some cases, large numbers of undergraduate students has been encouraging because it has allowed them to get their "feet wet" at an early stage in pursuit of a bachelor's and possibly a master's degree.

It is also apparent that, based on the data collected, some of the research grants have already terminated but extensions have been given for the purpose of completing final reports which draw conclusions based upon the research findings. At the time of this study, these termination dates range anywhere from one month to several years past their due date.
In cases where it is feasible, it would appear that such extensions could also involve new proposals which would allow
for continued funding past the technical date of termination. This would not only provide continuity of funding, but would also serve as a strategy for the utilization of remaining funds in any given year.

SUMMARY AND CONCLUSION

This study has examined the current status of NASA-JSC grants and contracts primarily to Historically Black Colleges and Universities. The one exception in this report is the University of Texas at El Paso, which falls under the MGR program. When one considers that there are some 110 HBCU's in the United States, the fact that only 26 hold grants and contracts from NASA as a whole is not, even

modestly speaking, excessive. This is due, in part, to the fact that not all HBCU's have engineering programs or departments, though it is safe to say that most have science programs or departments. Another factor that would impact upon the number of research grants and contracts awarded to HBCU's is the fact that their primary mission is teaching. Research is secondary or peripheral at best. Consequently, there is little incentive to seek research grants or contracts and such efforts are often not awarded by the school even when successful. These factors, then, serve as mitigating circumstances which affect research awards at HBCU's. Nevertheless, there is still a need to get more of these institutions involved as a result of Executive Order 12320 by the President of the United States in 1981.

Recommendations

Based upon the foregoing, the following recommendations are offered:

1.) the E.O. Office should use a tops-down as well as bottoms-up approach in publicizing the college programs in order to get maximum participation as well as cooperation from the technical divisions affecting and affected by the college programs,

2.) the E.O. Office should plan, structure and host a conference at JSC of the HBCU's regarding opportunities to receive grants and contracts from NASA,

3.) NASA in general and the JSC in particular should encourage more HBCU's to become research-oriented and provide release time for faculty with NASA grants,

4.) there is a need to develop a task force composed of JSC personnel to explore innovative ways and means to make the college programs successful and to assist the director of E.O. regarding related issues,

5.) it is suggested that the E.O. Office closely monitor research proposals as they travel through the NASA administrative process for approval or disapproval,

6.) it is recommended that all unsolicited proposals from HBCU's or related to the MGR program go directly to the E.O. Office before continuing through the proposal process,

7.) schools wth NASA grants and contracts should be notified in a systematic manner by the director of the E.O. Office as the termination date of their grant nears and encouraged to complete their final report on time (this is particularly true of Prairie View A&M University, which has had problems meeting the reporting schedule),

8.) in order for these programs to operate effectively and efficiently and carry out the numerous tasks and recommendations in this study, there is a need for a full-time staff person to successfully direct the program, and

9.) copies of both interim and final reports should be submitted to the E.O. Office as well as to the technical monitor to improve the coordination of final research products.

APPENDIX A

INSTITUTION/ ADDRESS	PI/PHONE	PROGRAM	GRANT TITLE	ORGANIZ. & BUYER	PI'S RACE/SEX	DATE OF GRANT	TERMINATION DATE	TECHNICAL MONITOR/PHONE
UNIVERSITY OF TEXAS AT EL PASO	DAN W. PATTERSON (915)747-5470	MGR	KNOWLEDGE BASED SYSTEM EMS	BE 31 RITZ, REX 38501	WHITE/M	09-01-86	01-01-88	MS. K, HEALY - EF 34776
TEXAS SOUTHERN UNIVERSITY NAG 9-174	R. WILSON 527-7011/ 527-7004/3	HBCU	MASS SPECTROMETRIC STUDY CERTAIN SPECIES	BE31 M. YOUNGBLOOD 38509	BLACK/M	09-01-86	09-01-87 (3 YR GRANT)	D. HENNINGER - SN12 35034
PRAIRIE VIEW A&M UNIVERSITY	JOHN OKYERE (409)857-3923 (409)857-3311	HBCU	STUDY CHARGED PARTICLE/RADIATION (LOW DOSE RADIATION AFFECTS ON VLSI)	BE21 MITZI BROYLES 38524	BLACK/M	03-30-84 5-85	06-01-87	C. HERMAN - EH3 38360
PRAIRIE VIEW A&M UNIVERSITY NAS 9-17406	WANG FA-CHANG (409)857-3311 (409)857-4510	HBCU	INTERACTIONS OF ATOMIC OXYGEN/ SPACECRAFT	BE21 KRISTEN ERICKSON 38513	ASIAN/P.I.	08-01-85	07-31-88	L. LEGER - ES5 38916 (RETURN 6/29)
ALABAMA A&M UNIVERSITY	PURUSHOPTAM KALE	HBCU	GENETIC EFFECTS OF HEAVY IONS IN DROSOPHILA	AA-7703 BEN GLISAN, JR.	ASIAN/P.I.	----	-----	STUART NACHTWEY- SD12 (37202) (RETURN 6/29)
PRAIRIE VIEW A&M UNIVERSITY	RONALD BOYD	HBCU	MARS SURFACED-BASED FACTORY PHASE I (PRE-LIMINARY DESIGN) PHASE II	NGT 44-003-900 (TRAINING GRANT)	BLACK/M	1/86	12/86	HEADQUARTERS

APPENDIX A (CONT.)

REPORTS REQUIRED	REPORTS RECEIVED	FUNDING AMOUNT	FUNDS DISBURSED	BALANCE	PROCUREMENT OFFICE(SCHOOL)/ PHONE	GRAD./UNDER GRAD. ASST.	RACE/SEX	LEVEL DEGREE	RENEWAL/COMMENTS
SEMI-ANNUALLY		96,929	22,658.50 5/30	74,270.50 5/30		2-1/2 / -	BLACK/M HISPANIC/F	MASTERS/ COMPUTER SCIENCE	STARTED 6 MONTHS LATE
PERIODIC REPORTS (SEMI-ANNUAL)	YES	110,382 (YEAR)	22,215.77 (6/22)	88,166.23	JOSEPH JONES/TSU	/2	BLACK/M	BS/ CHEMISTRY	SECOND YEAR PROPOSAL IN.
SEMI-ANNUAL FINAL REPORT DUE	NO	392,354	296,500.32 (6/23) 83,023.87 (5/30)	95,853.68 (6/23)	LARRY HOLDEN	4/1F 3M 20/	BLACK 18 BLACK/	MASTERS/E.E. 2 M.E./18 E.E.	POSSIBLE/PROBABLE (RECEIVED ONE GRADUATE THESIS OUT OF THREE)
QUARTERLY REPORTS/ FINAL	YES	194,861	50,000 5/30 WALTER 32277	144,861	LARRY HOLDEN	1/5	1 BLACK/M 4 BLACK/3M 1F 1 HISPANIC/ 1F	MASTERS/ENGIN. BS/ENGIN.	ENDS 7/88
						----	----	----	-----
YES	SEMI-ANNUALLY	125,000	NEW GRANT	NEW GRANT	JEWELL JOINER				
INTERIM REPORT OF FINAL REPORT FOR PHASE I	YES	57,000	50,000	7,000	LARRY HOLDEN	1/15	BLACK	INTERDISCIPLINARY	LEWIN WARREN (WROTE IN FEBRUARY) - NEVER RECEIVED REPLY

SOURCE: NASA JSC/LISTED EDUCATIONAL INSTITUTION

APPENDIX B

INSTITUTION/ ADDRESS	PI/PHONE	PROGRAM	GRANT TITLE	ORGANIZ. & BUYER	PI'S RACE/SEX	DATE OF GRANT	TERMINATION DATE	TECHNICAL MONITOR/PHONE
PRAIRIE VIEW A&M UNIVERSITY	RONALD BOYD (409)857-4023		A COMPREHENSIVE FIVE-TASK RESEARCH PROPOSAL	KRISTEN ERICKSON	BLACK/M			
PRAIRIE VIEW A&M UNIVERSITY	RONALD BOYD (409)857-4023	HBCU	SATURATED FLOW-BOILING HEAT TRANSFER W/FREON-11 IN TOP-HEATED HORIZONTAL COOLING CHANNELS: V	-----	BLACK/M	3/84	12/86	G. RANKIN - EC2 39131
PRAIRIE VIEW A&M UNIVERSITY	JOHN M. WOODS (409)857-2427	HBCU	EVALUATION OF PRIMARY LITHIUM BATTERY: TASK II	-----	WHITE/M	6/83	6/85	BOBBY BRAGG 39060 - EP5
PRAIRIE VIEW A&M UNIVERSITY	JOHN FULLER (409)857-3929	HBCU	LASER TARGET MOTION SIMULATOR: III	-----	BLACK/M	6/85	6/85 12/86	K. ERICKSON 38513
PRAIRIE VIEW A&M UNIVERSITY	ERVING EMMANEUL (NASA) 34280 34273	HBCU	MICROPROCESSOR BASED SOLAR VOLTAGE ARRAY: I	-----	BLACK/M	6/85	6/85	C. HERMAN EH13 38360
PRAIRIE VIEW A&M UNIVERSITY	DROPPED	HBCU	DISSOLVED IRON IN 204 PROPELLANT : (DROPPED) IV (FUNDS MOVED INTO OTHER TASKS)			DROPPED		H. BRASSEAUX EP 39026

APPENDIX B (CONT.)

REPORTS REQUIRED	REPORTS RECEIVED	FUNDING AMOUNT	FUNDS DISBURSED	BALANCE	PROCUREMENT OFFICE(SCHOOL)/ PHONE	GRAD./UNDER GRAD. ASST.	RACE/SEX	LEVEL DEGREE	RENEWAL/COMMENTS
YES	YES	$414,391	VOUCHER 9 6/15/87 403,948.35	10,442.65	----	-----	----	----	----------
PRELIMINARY REPORT SUBMITTED/ FINAL REPORT DUE	YES	93,766	-----	-----	LARRY HOLDEN (409)857-3311	2/0 (NOT NASA FUNDED)	BLACK/M	MECH. ENG. MASTERS	SEEKING ADDITIONAL FUNDING EXTENSION GIVEN
ONE FINAL REPORT STATUS REPORTS	YES	105,616	-----	-----	------	- / -	- / -	- / -	PROBABLY WILL SUBMIT NEW PROPOSAL
FINAL REPORT SENT IN	YES	93,080	-----	-----	------	2-1/2 / 10	1 IRANIAN/M 1-1/2BLACK/M 10 BLACK/M	MASTERS/ ENG. BS/ENG.	TALKING W/JANE GRADY. PROPOSAL IN TO CONTINUE (INFORMAL)
PRELIMINARY FINAL REPORT TO FULLER DR. ORKEYE WILL SUBMIT FINAL REPORT	YES	55,718	-----	-----	L. HOLDEN	-/6 (PAID)	BLACK/3M 3F	BSEE	NO EFFORT TO RENEGOTIATE.
DROPPED	DROPPED	66,211	----	----	------				
	TOTAL	$414,391	$403,948.35	$10,442.65					

REFERENCES

1. Mfanya Donald L. Tryman, Affirmative Action As Organization Development At The Johnson Space Center, Houston, NASA-Johnson Space Center, July 1986, p. 44.

2. Ibid.,p. 39.

3. Office of the Chief Scientist, "National Trends That May Impact NASA," Washington, D.C.: NASA Headquarters, p. 3.

4. Proclamation of the President of the United States, "National Historically Black Colleges Week, 1986," Document, Washington, D.C., 1986.

5. NASA-HBCU Forum '86, Westin Peachtree Plaza Hotel, Atlanta: April 20-22, 1986, p. 1.

6. Office of the Chief Scientist, NASA Headquarters, "Historically Black Colleges and Universities," October 3, 1986, p. 1.

7. Frank B. McDonald, "Letter to NASA Agencies," September 5, 1986, p. 1.

8. Ibid., Enclosure 1.

9. Ibid., p. 1.

10. Minutes of the NASA EO Council Meeting, Houston: NASA-Johnson Space Center, October 7, 19886, p. 11.

11. Frank B. McDonald, op. cit., p. 2.

12. Minutes, op. cit., p. 11.

13. Jet Magazine, "10 Richest Black Colleges," Vol. 72, No. 12, June 15, 1987, p. 14.

Bibliography

McDonald, Frank B., "Letter to NASA Agencies," Washington, D.C.: Office of the Chief Scientist, NASA Headquarters, September 5, 1986.

Minutes of the NASA EO Council Meeting, Houston: NASA-Johnson Space Center, October 7, 1986, p. 11.

NASA-HBCU Forum '86, Atlanta: Westin Peachtree Plaza, April 20-22, 1986.

Office of the Chief Scientist, "National Trends That May Impact NASA," Washington, D.C.: NASA Headquarters, October 1, 1986.

Office of the Chief Scientist, "Historically Black Colleges and Universities," Washington, D.C.: NASA Headquarters, October 3, 1986.

Proclamation of the President of the United States, "National Historically Black Colleges Week, 1986," Washington, D.C.: Document, 1986.

Tryman, Mfanya Donald L., Affirmative Acton As Organization Development At The Johnson Space Center, Houston: NASA-Johnson Space Center, July 1986.

THE DISTRIBUTION OF EVAPORITIC WEATHERING PRODUCTS ON ANTARCTIC METEORITES

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Michael A. Velbel, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	Michigan State University Dept. of Geological Sciences East Lansing, MI 48824-1115
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
Branch:	Planetary Materials
JSC Colleague:	James L. Gooding, Ph.D.
Date:	August 7, 1987
Contract:	NGT 44-001-800

ABSTRACT

White evaporite deposits of terrestrial origin occur on some 5% of Antarctic meteorites. The few previous studies, and new mineralogical analyses, indicate that the deposits are predominantly carbonates and/or sulfates of magnesium. The distribution of white evaporitic salt deposits differs among different meteorite compositional groups and weathering categories of Antarctic meteorites. Salts occur with unusual frequency on carbonaceous chondrites, and are especially common in carbonaceous chondrites of weathering categories A and B. Among achondrites, weathering categories A and A/B show the most examples of salt weathering. Unlike carbonaceous chondrites and achondrites, most salt-bearing ordinary (H and L) chondrites are from rustier meteorites of weathering categories B, and, to a lesser degree, B/C and C. LL chondrites are conspicuous by their complete lack of any salt-weathering product.

Almost two-thirds of all evaporite-bearing meteorites belong to weathering categories A, A/B, and B. Where chemical and/or mineralogical data are available, there is a persistent suggestion that evaporite formation is accompanied by elemental redistribution from meteorite interiors. Meteorites of weathering categories B, A/B, and even A may have experienced significant element redistribution and/or contamination as a result of terrestrial exposure. Evaporite formation during terrestrial weathering cannot be neglected in geochemical, cosmochemical, and mineralogical studies of Antarctic meteorites. A lower-case "e" should be added to the weathering classification of evaporite-bearing Antarctic meteorites, to inform meteorite scientists of the presence of evaporite deposits and their associated compositional effects.

INTRODUCTION

Terrestrial weathering interferes with the use of meteorite mineralogy and chemistry as windows into the pre-terrestrial history of the meteorites, their parent bodies, and the early solar system, because weathering transforms original minerals to alteration products and redistributes chemical elements. Nowhere is this phenomenon of greater concern to meteorite scientists than in Antarctica. More than 7000 samples, representing several thousand meteorites, have been collected in Antarctica in the past two decades (Lipschutz and Cassidy, 1986). However, none of these are fresh falls.

Although weathering most likely proceeds only slowly in the cold, dry Antarctic environment, Antarctic meteorites have been exposed to terrestrial conditions for tens of thousands to hundreds of thousands of years (e.g., Nishiizumi, 1986; Schultz, 1986a). Thus, ample opportunity exists for small increments of weathering to accumulate into measurable effects. Characterization of terrestrial weathering effects is therefore crucial to studies of Antarctic meteorites (e.g., Gooding, 1986a,b), in order to prevent terrestrial alterations from being misinterpreted as primary compositional attributes of the meteorites.

The present procedure for classifying the weathering of Antarctic meteorites is based on the amount of rust visible to the unaided eye; weathering categories "A", "B", and "C" indicate, respectively, "minor", "moderate", and "severe" rustiness. However, other forms of terrestrial alteration, which do not involve the formation of rust, are also known. The most obvious and widely noted of these "non-rusty" alterations is the appearance of white powders or efflorescences on the surfaces of meteorites. Although relatively rare, occurring on less than 5% (fewer than 100) of the samples, the white deposits are nevertheless the second most widely recognized form of terrestrial alteration, after the formation of rust. The white deposits clearly indicate some combination of post-fall elemental redistribution and/or terrestrial contamination. Characterizing these deposits is therefore essential to understanding the nature and magnitude of weathering effects in Antarctic meteorites.

The purpose of this project is to twofold: 1) To investigate the mineralogy, chemistry, and origin of white efflorescences on the Lewis Cliff, Antarctica, 85320 (H5) chondrite (hereinafter referred to as LEW 85320). The unusual abundance of material available from LEW 85320 provided a rare opportunity to employ a variety of analytical methods. 2) To examine the occurrence of evaporite deposits on Antarctic meteorites as a function of meteorite compositional group and weathering category.

PREVIOUS WORK

Previous work on the white powdery deposits has been limited to a handful of descriptive mineralogical studies on eight meteorites (Yabuki *et al.*, 1976; Marvin, 1980; Marvin and Motylewski, 1980), which demonstrated that the deposits consist of various evaporite salts of Mg. Yabuki *et al.* (1976) identified the white material on Y 74371 as nesquehonite ($MgCO_3 \cdot 3H_2O$). Marvin (1980; Marvin

and Motylewski, 1980) identified several evaporite minerals, including hydromagnesite, epsomite, leonhardite, gypsum, and nesquehonite, on seven Allan Hills meteorites, which included two H chondrites, two L chondrites, two carbonaceous chondrites, and one ureilite.

Marvin (1980) noted that "the sources of the elements and processes leading to formation of the evaporites are problematical", and opinion on the origin of the white deposits remains divided. In their discussion of the origin of the nesquehonite on Y 74371, Yabuki *et al.* (1976) favored possible external sources of magnesium, such as sea water, spring water, saline lake waters, volcanic emanations, alkaline hot springs, or hydrothermal solutions, despite the obvious difficulty in explaining how a meteorite which had been exposed to such an environment could have been found in Antarctic ice in association with unaltered meteorites. In contrast, additional work led Marvin and Motylewski (1980) to "conclude that Mg, S, and minor elements are leached from the meteorite. The source of carbon is probably atmospheric CO_2." The origin of these deposits is still uncertain.

Marvin (1980) commented that "the deposits are not restricted to any one variety of meteorite." There was no apparent correlation of evaporite occurrence with iron-oxide staining; only two of the salt-bearing meteorites exhibited any significant degree of iron-oxide staining (weathering categories B and C). Previous work involved too small a number of samples to permit any inferences regarding relationships between the occurrence of salts, and either meteorite composition, or meteorite rusting, in Antarctic meteorites.

PART I: EVAPORITES ON LEW 85320

METHODS

LEW 85320 is a large (>110 kg), moderately rusty (weathering category B) H5 chondrite. Two samples were examined in this study. Sample 40 consists of millimeter-scale chips (aggregates) of pure white powdery material, scraped from surficial accumulations which were observed in the field at the time the sample was recovered. Sample 22 consists of surficial scrapings, whose "salt and pepper" appearance suggests a mixture of white evaporite deposit and dark brown or black fusion crust. The white material of sample 22 appeared upon return of the sample to JSC, during controlled laboratory drying.

Mineralogical characterization utilized X-ray diffraction (XRD) and scanning electron microscopy with energy-dispersive X-ray spectrometry (SEM/EDS). For XRD analysis, a portion of each sample was ground in acetone on a glass microscope slide, using a glass stirring rod, and the resultant slurry was allowed to dry on the slide. These preparations were then scanned from 5^{o} to $50\text{-}60^{o}$ two-theta in a Phillips X-ray diffractometer with Cu radiation, using the following instrumental settings; 40kV, 20 mA, 1^{o} divergence and receiving slits, scanning rate of 1^{o} two-theta per minute, chart recorder speed of 10mm per minute, time constant of 1 second, and chart scale settings appropriate to the amplitudes of the most intense peaks.

Samples for SEM/EDS analysis were mounted on graphite stubs, using colloidal graphite as the adhesive, and a conductive coating of carbon was applied with a vacuum evaporator. Samples were examined using a JEOL-35CF SEM equipped with a PGT4000T analyzer, operating at 20kV.

RESULTS

XRD results are summarized in Table 1. Sample 40, the pure white powdery aggregate, produced an excellent XRD pattern (Table 1), with twenty-six peaks discernable above background between 2 and 60 degrees two-theta. Eleven of these correspond to the peaks for pure synthetic nesquehonite ($MgCO_3 \cdot 3H_2O$) reported by Ming and Franklin (1985). All but one have been reported previously in natural samples of nesquehonite (Yabuki *et al.*, 1976 and Ming and Franklin, 1985, and references therein; ASTM 20-669), although some of them may be attributable to the presence of small amounts of other hydrous magnesium carbonates, such as hydromagnesite [$Mg_5(CO_3)_4(OH)_2 \cdot 4H_2O$], ASTM 25-513; see also Ming and Franklin, 1985) or barringtonite [$MgCO_3 \cdot 2H_2O$], ASTM 18-768. The one anomalous peak, at 1.03 nm, is quite weak relative to the nesquehonite peaks. It probably represents trace quantities of another hydrous magnesium carbonate, the hexahydrate, "hydrophase II" [$Mg_5(CO_3)_4(OH)_2 \cdot 6H_2O$] (Ming, 1981), whose two principal XRD peaks (1.03 and 5.79 nm) are an excellent match to the 1.03 and 5.82 nm peaks in the XRD pattern of LEW 85320,40. If the weak 1.03 nm peak represents of Ming's (1981) hydrophase II, then LEW 85320,40 is the first reported natural occurrence of the hexahydrate. All peaks are attributable to nesquehonite and traces of other hydrous magnesium carbonates, suggesting that sample 40 is pure carbonate.

TABLE 1 - XRD DATA FOR METEORITIC AND TERRESTRIAL NESQUEHONITE

PRESENT WORK LEW85320 ,22		,40		Ming & Franklin Synthetic		Yabuki et al. Yamato 74371		Synth.		ASTM 1-1030	
d(nm)	I	d(nm)	I	d(nm)	I	d(nm)	I	d(nm)	I	d(nm)	I
		1.03	5								
0.651	**22**	**0.651**	**>100**	0.651	10	0.652	VS	0.646	VS	0.650	10
						0.607	vw	0.606	vvw		
		0.582	5							0.580	0.8
		0.490	2			0.491	w	0.490	mw	0.496	0.8
						0.458	vw				
		0.446	2								
		0.413	4			0.415	vw	0.417	vw		
0.385	**26**	**0.386**	**93**	0.386	9	0.386	VS	0.385	VS	0.386	8
0.359	**8**	**0.357**	**16**	0.356	3	0.359	m	0.357	m	0.358	1.6
						0.335	vvw				
0.323	**11**	**0.323**	**12**	0.322	2	0.323	m	0.323	m	0.323	1.6
0.303	**9**	**0.303**	**30**	0.302	3	0.304	S	0.305	ms	0.302	2.4
		0.298	6			0.298	w	0.298	w		
						0.288	vw				
						0.283	m				
0.278	**15**	**0.278**	**11**	0.277	2	0.278	m	0.278	m	0.277	1.6
0.262	**10**	**0.262**	**13**	0.261	2	0.262	S	0.262	S	0.261	4.8
0.252	**12**	**0.251**	**21**	0.250	4	0.251	m			0.251	4
								0.249	m		
		0.247	4			0.247	vw				
		0.234	2			0.234	m	0.234	w	0.235	0.8
						0.231	vw				
						0.220	vvw				
						0.218	w				
		0.217	**9**	0.217	1	0.217	w	0.217	w	0.217	1.6
						0.216	vw				
						0.212	vw				
		0.202	**8**	0.201	1	0.202	mw	0.201	mw	0.201	0.8
						0.201	vw				
0.193	**8**	**0.192**	**19**	0.192	3	0.193	m	0.193	m	0.192	3.2

Two-theta > 48 not reported.

BOLD FACE = synthetic nesquehonite Ming & Franklin (1985)
underline = reagent grade hydromagnesite and/or synthetic magnesium hydroxycarbonate hexahydrate (Ming, 1981)
regular face = peak unknown from well-characterized synthetic phase.

Sample 22 yielded relatively low peak intensities, most likely because of poor crystallinity of the salt (see SEM results, below), although the presence of glassy fusion crust in the sample may also have played a role. Sample 22 yielded two strong (0.651 and 0.385 nm) peaks, and some half-dozen very weak peaks, barely above background (Table 1). Despite the poor quality of the XRD pattern for this sample, all peaks which exist correspond to peaks reported from natural (Yabuki *et al.*, 1976) and synthetic (Yabuki *et al.*, 1976; Ming and Franklin, 1985) nesquehonite.

Scanning electron microscopy of aggregates from sample 40 reveals that minerals of the evaporite deposits possess two distinctive morphologies. The bulk of the material exhibits an elongate, lath-like prismatic habit. Individual laths are generally 30-60 microns long, and 5-10 microns wide. Prism faces often exhibit striae parallel to the direction of elongation. Prism edges are sharp and distinct, although the ends of the prisms are commonly somewhat rounded. The rounding of the ends of the laths may be due to small amounts of dissolution-reprecipitation taking place in the presence of water sometime after the crystals formed. The laths are arranged in a felted or jackstraw fabric. EDS analysis reveals Mg to be the only detectable metal cation. The elongate habit is characteristic of both natural (Yabuki *et al.*, 1976) and synthetic (Ming and Franklin, 1985) nesquehonite.

The second characteristic morphology is a fragile wispy form, consisting of thin (submicron) gently curved sheets. The distribution ranges from individual wisps superposed upon the prism faces and rounded edges of the laths, through small botryoidal clusters of anastomosing wisps, to large (100 micron) clumps of intricate porous intergrowths, which appear to have formed by coalescence of smaller clusters. Partial or complete encrustation of laths is also occasionally observed. EDS analysis again reveals Mg to be the only detectable metal cation. The clumps of wispy magnesian material are morphologically very similar to synthetic "hydromagnesite hydrophases" prepared by Ming (1981). Although striking in appearance, the wisps are volumetrically insignificant compared to the laths. The wispy material may be the small amount of hydromagnesite (and/or the hexahydrate "hydrophase II" of Ming, 1981) responsible for many of the non-nesquehonite peaks in the XRD results. The superposition relationship with the prismatic and rounded surfaces of the nesquehonite laths indicates that both nesquehonite precipitation and rounding of the nesquehonite crystals preceded the formation of the hydromagnesite.

Sample 22 is similar to sample 40, except that the bulk of the Mg-rich material shows no crystal faces, and appears fractured in places. This material volatilizes under the electron beam exactly as nesquehonite does, and, because it makes up the bulk of the material under the SEM, is inferred to be the crystalline phase responsible for the nesquehonite peaks in the XRD pattern for this sample. Thin superposed deposits of wispy material are identical to those of sample 40.

DISCUSSION

Conditions of nesquehonite formation on LEW 85320 can only be speculated upon, but morphological and experimental results both suggest that nesquehonite may have formed under Antarctic conditions. Ming and Franklin (1985) synthesized nesquehonite at temperatures as low as 4°C. Schultz (1986b) experimentally determined that temperatures in the interior (2 cm depth) of a sample of the Allende (CV3) meteorite exposed to Antarctic conditions reached 5°C on wind-free days during austral summer, despite the fact that the air temperature never exceeded -11°C over the study period. Thus, insolation heating of meteorites in Antarctica is sufficient to 1) raise the temperture of meteorite interiors above the freezing point of pure water, to depths of several centimeters, and 2) to attain temperatures at which nesquehonite may form directly from the resultant solutions. Therefore, the observed mineralogy and morphology of Antarctic meteoritic nesquehonite, and present knowledge of nesquehonite stability, are both consistent with direct formation of nesquehonite under Antarctic conditions.

Ordinary chondrites consist predominantly of olivine and pyroxene, with varying abundances of Fe-Ni metal and sulfides. The presence of hydrous magnesium carbonates as alteration products implies that water and carbon were also available to the weathering system. Gooding's (1981) mineralogical results and Schultz's (1986b) experimental data indicate that liquid water is at least occasionally available to Antarctic meteorites. If we assume that the Mg for the nesquehonite was derived from the meteorite itself (e.g., Marvin and Motylewski, 1980), a reaction involving the ferromagnesian silicate minerals of the ordinary chondrite is required to release the Mg. A single reaction, using olivine as the mineral reactant, can be written which accounts for all these requirements:

$$(Fe_xMg_{1-x})_2SiO_4 + 8-5xH_2O + 2-2xCO_2 + 0.5xO_2 \Longrightarrow$$

$$2-2x(MgCO_3 \cdot 3H_2O) + 2xFeOOH + H_4SiO_4$$

where x is the mole fraction fayalite in the meteoritic olivine. Carbon dioxide combines with water to form carbonic acid, which dissociates to give a hydrogen ion and bicarbonate. The hydrogen ion drives the olivine hydrolysis reaction, liberating Mg. Magnesium reacts with bicarbonate and water to form nesquehonite. Oxygen oxidizes the Fe, which reacts with water to form goethite, and silicon is released to solution, from which it can be precipitated as some form of silica.

The hypothesized internal source for Mg requires mass transfer from meteoritic silicates to the surficial evaporites, but formation of evaporites in the observed amounts would have no detectable effect on the MgO content of the bulk meteorite. Approximately one gram of evaporite was recovered from the 110kg LEW 85320 H chondrite. Pure nesquehonite is about 28% by weight MgO; thus, the recovered mass of evaporite contains 0.28g of MgO. If LEW 85320 has the average MgO content of Antarctic H chondrites (21.82 wt%; Fulton and Rhodes, 1984), the total MgO inventory of the meteorite is 24kg. Therefore, the amount of MgO in the evaporite deposit is about 10^{-5} of the meteorite's MgO inventory, an amount so small that its removal from the meteorite interior would not even affect the last decimal place in MgO analyses reported to four significant figures.

The olivine-nesquehonite reaction illustrates how the terrestrial atmosphere and hydrosphere can react with meteoritic ferromagnesian silicates to form all of the widely observed terrestrial alteration products of Antarctic meteorites, *without* resorting to external sources for any lithophile cations. Hydrous magnesium carbonates result directly from this reaction. The iron oxy-hydroxide, goethite, contributes to the abundant metallic rust (e.g., Gooding, 1986a) which is the most conspicuous alteration phenomenon of Antarctic meteorites, and the silica is available to contribute to "sialic rust", which was analyzed and discussed by Gooding (1986a).

PART II: DISTRIBUTION OF EVAPORITES ON ANTARCTIC METEORITES

METHODS

If the occurrence of salt were independent of meteorite composition and controlled exclusively by conditions in the weathering environment, we might expect that the number of occurrences of evaporite deposits on meteorites of any individual class would be proportional to the relative abundance of individuals of that class among all Antarctic meteorites. In other words, if 1) meteorite class Z constitutes 20% of all Antarctic meteorites, 2) all meteorites are equally likely to encounter evaporite-forming environments (there's no giant magnet selectively drawing some classes of meteorites towards salt-forming environments), and 3) all meteorites are equally likely to form evaporite minerals if subjected to salt-forming environments, then meteorite class Z should constitute 20% of the reports of salt-bearing meteorites. Thus, if occurrence of salt were independent of meteorite composition (as Marvin's, 1980, preliminary data suggested), all meteorites would be equally vulnerable to salt-formation, and occurrences of evaporite salts would be apportioned among all groups of meteorites in proportion to the groups' relative abundances.

To test the hypotheses that occurrence of evaporite salts is independent of meteorite composition or weathering class, a census of evaporite occurrences on Victoria Land, Antarctica, meteorites was taken, using data from several sources: 1) Previous work on salt weathering of Antarctic meteorites (Marvin, 1980); 2) Preliminary descriptions of Antarctic meteorites (from the Antarctic Meteorite Newsletter [AMN]); and 3) Examination of photographs in the Antarctic Meteorite Collection at the NASA/Johnson Space Center. Results were tabulated both on the basis of individual unpaired samples, and compensated for pairing. Pairing was accounted for by counting all samples from a given pairing (AMN) as one meteorite, and assigning it to the weathering category of the most highly weathered individual sample. The results of this census were then compared with data on Antarctic meteorite abundances (uncorrected and corrected for pairing) provided by R. Score (written communication, June 30, 1987, and personal communication, July 15, 1987). This census of white evaporite deposits covers all U.S. Antarctic meteorites described and/or photographed before mid-July, 1987.

RESULTS

Results of this census are summarized in Table 2. (A complete list of all evaporite-bearing meteorites is available from the author upon request.) Ratios of evaporite-bearing-to-total meteorites in each category are shown as percentages in Table 3.

This study reveals that the occurrence of evaporite salts is definitely influenced by meteorite composition, for certain groups of meteorites. Evaporite occurrence is also correlated with rustiness, but salt coexists with different weathering classes among different compositional groups of meteorites. The influence of environmental factors can be seen in the data for paired meteorites.

Major features of the census results indicate:

1) A strong compositional influence is suggested for carbonaceous chondrites. Although carbonaceous chondrites make up less than two per cent of the total number of Antarctic meteorites, they constitute nearly one-fifth (19.4%) of the described occurrences of white evaporite deposits. A much larger number of carbonaceous chondrites possess salts than would be expected if all classes of meteorites were equally vulnerable to salt formation. Composition appears to play a major role in determining the distribution of evaporite salts in carbonaceous chondrites.

2) A compositional influence is also suggested for the E chondrites, LL chondrites, and possibly the achondrites. E chondrites make up less than one percent of the total population of Antarctic meteorites, yet they constitute more than 4% of the reported occurrences of white deposits. LL chondrites make up 5% of the total population, yet no occurrences of salt-bearing LL chondrites have been reported. Achondrites constitute less than 4% of the total population, yet make up more than 10% of the evaporite-bearing meteorites.

3) For some groups of meteorites (L and H chondrites, and irons/stony-irons), the proportion of meteorites of that group among the evaporite-bearing meteorites is similar to (within a factor of three of) their relative abundance among the total population of Victoria Land meteorites, suggesting that there is no significant difference in the susceptibility of these meteorites to evaporite formation.

4) Composition is not, however, the sole determining factor. There are several paired meteorites for which different individual samples have different weathering classifications. In the case of one carbonaceous chondrite with nearly a dozen evaporite-bearing pieces, weathering classifications for individual samples ranged

TABLE 2 - SALT WEATHERING STATISTICS

	WEATHERING CATEGORY					SALT-BEARING		ALL ANTARCTIC[1]	
GROUP	A	A/B	B	B/C	C	#	%	#	%
C	6	2	5	1		14	19.4	26	1.9
LL						0		66	4.8
L	1	1	11	4	3	20	27.8	415	30.4
H	1		11	9	4	25	34.7	782	57.2
E		1		1	1	3	4.2	9	0.7
ACHONDR.	3	3	1		1	8	11.1	49	3.6
IRONS & STONY IRONS						2	2.8	20	1.5
TOTAL[2]						72	100	1367	100
TOTAL[3]	11	7	28	15	9	70			
%	15.7	10.0	40.0	21.4	12.9	100			

[1]Includes 78 classified meteorites (irons and stones) which were not assigned a weathering category.

[2]Includes two salt-bearing irons.

[3]Does not include the 2 salt-bearing irons, because irons are not assigned a weathering category.

TABLE 3 - RATIOS OF EVAPORITE-BEARING/TOTAL FOR EACH GROUP AND WEATHERING CATEGORY (%)

	WEATHERING CATEGORY					PROPORTION OF
GROUP	A	A/B	B	B/C	C	ENTIRE GROUP
C	85.7	25.0	55.6	50.0		53.8
LL						
L	2.8	1.3	5.9	6.9	5.9	4.9
H	2.4	0	4.0	5.5	2.1	3.4
E	0	50.0	0	100.	50.0	33.3
ACHONDR.	27.3	20.0	7.1	0	50.0	16.7

from A to B/C! There are also several reports of paired meteorites in which only one individual possesses white deposits. Differences in the weathering behaviors of paired meteorites cannot be due to differences in meteorite composition; therefore, some external environmental factor must account for non-identical weathering of such pairs.

5) Almost two-thirds (65.7%) of the evaporite-bearing stony meteorites are classified as weathering categories A, A/B, or B. Most of the evaporite-bearing but less-rusted meteorites are carbonaceous chondrites of weathering categories A and B, and achondrites of weathering categories A and A/B. Significant evaporite formation is correlated with the earliest stages of rusting (weathering categories A and B) in carbonaceous chondrites. Sixth-sevenths (85.7%) of the carbonaceous chondrites of weathering category A possess salt. The absence of evaporite-bearing carbonaceous chondrites from weathering category C does not indicate negative correlation between salt and rust in these meteorites, but merely the fact that there are no carbonaceous chondrites in this weathering category.

6) Most evaporite-bearing ordinary chondrites are assigned to weathering category B. Evaporite formation is infrequently associated with the earliest stages of rusting of ordinary chondrites (3 evaporite-bearing meteorites out of 223 H and L chondrites in weathering categories A and A/B), but dominates in weathering categories B, B/C and C, in these meteorites (42/916).

DISCUSSION

The susceptibility of an individual meteorite to evaporite formation is a function of meteorite group. For example, carbonaceous chondrites make up a much larger proportion of the salt-bearing population than they make up of the overall population of Antarctic meteorites, and more than half of all C chondrites possess white deposits. Apparently, the high volatile content of carbonaceous chondrites is conducive to the formation of salts, possibly because these meteorites are capable of supplying their own oxidizable carbon and/or sulfur, which can combine with water to supply acid for silicate hydrolysis reactions and acid anions for evaporite formation.

As noted above, the formation of evaporite efflorescences requires either remobilization of elements from within the sample or addition of elements from outside. It is thus evident from the large number of evaporite occurrences on "unweathered" and "moderately

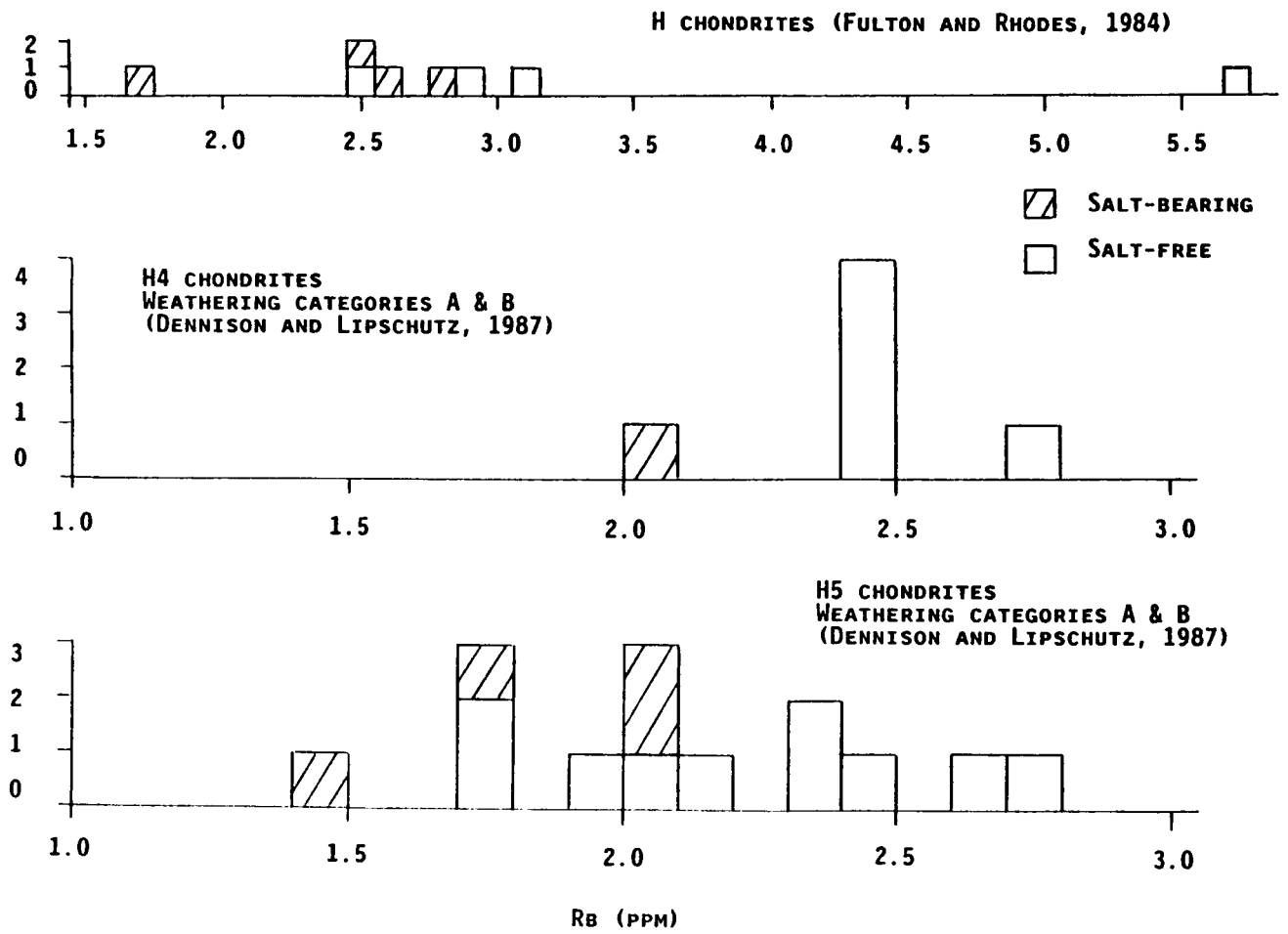
weathered" meteorites that, despite assertions to the contrary (e.g., Dennison and Lipschutz, 1987), even meteorites of weathering categories A, A/B, and B may have experienced significant element redistribution and/or contamination as a result of terrestrial exposure.

These findings may require re-evaluation of earlier conclusions regarding the influence of Antarctic meteorite weathering on elemental composition. For example, evaporite-bearing meteorites of weathering categories A and B were included in several previous compositional studies of Antarctic ordinary chondrites (Dennison and Lipschutz, 1987; Fulton and Rhodes, 1984). These previous workers concluded that meteorites of weathering categories A, A/B, and B were "compositionally uncompromised" by weathering. However, all but one of the ostensibly "uncompromised" meteorites which possess salt deposits have interiors which contain below-average abundances of Rb (Figure 1). Similarly, Kallemeyn (1986) noted that two other salt-bearing chondrites show some depletion of Ca. There are no corresponding reports of Mg depletion from the interiors of evaporite-bearing meteorites, but, as noted above, small losses of Mg most likely would go unnoticed due to the large absolute amount of Mg in most stony meteorites.

Just as the interiors of many evaporite-bearing meteorites appear to be depleted in alkali and alkaline-earth elements, exteriors of evaporite-bearing meteorites appear to be enriched in these elements. The exterior of evaporite-bearing ureilite ALHA77257 (Marvin, 1980) is significantly enriched in Rb and Cs (Biswas *et al.*, 1980; Lipschutz, 1982).

Evaporite-bearing meteorites seem to have interiors depleted in Rb (and Ca?). The presence of alkaline-earth salts, and observed enrichment of meteorite surfaces in alkalis, require a source of alkali and alkaline earth elements. External derivation and contamination of exterior surfaces (Biswas *et al.*, 1980; Lipschutz, 1982) is an unsatisfactory hypothesis, because it would not account for below-average Rb contents in the interiors of salt-bearing meteorites. Parsimony favors a meteoritic origin for the alkaline earth elements of the evaporite deposits. Leaching and transfer of alkalis and alkaline earths from meteorite interiors, followed by reprecipitation of those elements as salt deposits on meteorite exteriors, would qualitatively satisfy material balance requirements, and account for all these observations without resorting to external sources for these elements. This explanation is consistent with the inferences of Marvin and Motylewski (1980) and Gooding (1981).

FIGURE 1 - RB IN ANTARCTIC H CHONDRITES



SUMMARY, AND IMPLICATIONS FOR CLASSIFICATION OF METEORITE WEATHERING

The distribution of evaporite deposits varies with meteorite compositional group and degree of rustiness. Proportions of salt-bearing meteorites vary among compositional groups. There is no one-to-one correspondence between evaporites and rust; salts coexist with all degrees of rustiness in all groups of meteorites. Within compositional groups, certain weathering categories contain higher proportions and abundances of evaporite-bearing meteorites, but this association of salt with rust varies from group to group.

Interiors of evaporite-bearing meteorites of weathering categories A, A/B, and B have lower Rb concentrations than interiors of equally rusty salt-free meteorites. This suggests three things: 1) Dennison and Lipschutz (1987) are correct in suggesting that leaching is the principal mode of trace- (and major-?) element mobilization during Antarctic meteorite weathering. However, 2) contrary to Dennison and Lipschutz's (1987) and Fulton and Rhodes' (1984) conclusions, meteorites of weathering categories A and B also exhibit leaching effects. 3) Leaching of elements from the interiors of chondrites of weathering categories A and B is associated with the formation of visible evaporite deposits (Figure 1; Table 1). Thus, evaporite efflorescences serve as a warning that, despite minimal rusting, the sample in question is compositionally compromised.

Given the fact that evaporite occurrence is not directly correlated with rustiness, and that evaporite occurrence indicates element redistribution, it is apparent that the present description of weathering categories is inadequate to describe potentially compromised samples. Until a more objective "weatherometer" (Gooding, 1986b) is developed, I propose an interim solution. A lower-case "e" can be appended to the weathering category entry for all meteorites possessing visible evaporite salts or other efflorescences. This follows the same approach as that used to add information on shock metamorphism to the compositional classification of meteorite groups (Dodd, 1981). Thus, for example, the weathering category entry for ALHA77257 would be "Ae", indicating both the minor degree of rustiness, and the presence of white evaporite deposits. This proposal involves minimal modification of existing data-bases, yet it provides a warning flag to meteorite investigators, informing them of the presence of a potentially significant terrestrial weathering feature which is not included in the present weathering classification.

CONCLUSIONS

The distribution of white evaporitic deposits differs among different meteorite compositional groups and weathering categories in Antarctic meteorites. Salts occur with unusual frequency on carbonaceous chondrites, and are especially common in carbonaceous chondrites of weathering categories A and B. Among achondrites, weathering categories A and A/B show the most examples of salt weathering. Unlike carbonaceous chondrites and achondrites, most salt-bearing ordinary (H and L) chondrites are from rustier meteorites of weathering categories B, and, to a lesser degree, B/C and C. LL chondrites are conspicuous by their complete lack of salt-weathering products.

Almost two-thirds of all evaporite-bearing meteorites belong to weathering categories A, A/B, and B. Where chemical and/or mineralogical data are available, there is a persistent suggestion that evaporite formation is accompanied by elemental redistribution from meteorite interiors. The distribution of evaporite salts militates against the hypothesis that meteorites of weathering categories A, A/B, and B are "compositionally uncompromised" by weathering. Evaporite formation during terrestrial weathering is a significant consideration in geochemical, cosmochemical, and mineralogical studies of Antarctic meteorites.

In order to provide meteorite investigators adequate warning of possible salt-related element redistribution and/or terrestrial contamination, a lower-case "e" can be appended to the weathering category entry for all Antarctic meteorites possessing visible evaporite salts or other efflorescences. This proposal involves minimal modification of existing data-bases, yet it provides a warning flag to meteorite investigators, informing them of the presence of a potentially significant terrestrial weathering feature which is not included in the present weathering classification.

REFERENCES

Biswas, S.; Ngo, H.T.; and Lipschutz, M.E.: Trace element contents of selected Antarctic meteorites, I: Weathering effects and ALHA-77005, A77257, A77278 and A77299. Zeitschr. fur Naturforschung, vol. 35a, 1980, pp. 191-196.

Dennison, J.E.; and Lipschutz, M.E.: Chemical studies of H chondrites. II: Weathering effects in the Victoria Land, Antarctic population and comparison of two Antarctic populations with non-Antarctic falls. Geochim. Comsochim. Acta, vol. 51, 1987, pp. 741-754.

Dodd, R.T.: Meteorites, a Petrologic-Chemical Synthesis. Cambridge University Press, Cambridge, 1981.

Fulton, C.R.; and Rhodes, J.M.: The chemistry and origin of the ordinary chondrites: Implications from refractory-lithophile and siderophile elements. Proc. 14th Lunar Planet. Sci. Conf., Part 2, Jour. Geophys. Res., vol. 89, Supplement, 1984, pp. B543-B558.

Gooding, J.L.: Mineralogical aspects of terrestrial weathering effects in chondrites from Allan Hills, Antarctica. Proc. 12th Lunar Planet. Sci. Conf., 1981, pp. 1105-1122.

Gooding, J.L.: Clay-mineraloid weathering products in Antarctic meteorites. Geochim. Cosmochim. Acta, vol. 50, 1986a, pp. 2215-2223.

Gooding, J.L.: Weathering of stony meteorites in Antarctica. In International Workshop on Antarctic Meteorites, J.O. Annexstad, L. Schultz, and H. Wanke, eds., Lunar and Planetary Institute Tech. Report 86-01, 1986b, pp. 48-54.

Kallemeyn, G.W.: A compositional study of the ungrouped Y793495 carbonaceous chondrite and several CM chondrites from Antarctica. Meteoritics, vol. 21, 1986, p. 410.

Lipschutz, M.E.: Weathering effects in Antarctic meteorites. In Catalog of Meteorites from Victoria Land, Antarctica, 1978-1980, U.B. Marvin and B. Mason, eds., Smithsonian Contrib. Earth Sci., no. 24, 1982, pp. 67-69.

Lipschutz, M.E.; and Cassidy, W.A.: Antarctic meteorites: A progress report. EOS (Trans. Amer. Geophys. Union), vol. 67, 1986, pp. 1339-1341.

Marvin, U.B.: Magnesium carbonate and magnesium sulfate deposits on antarctic meteorites. Antarct. Jour. U.S., vol. 15, 1980, pp. 54-55.

Marvin, U.B.; and Motylewski, K.: Mg-carbonates and-sulfates on Antarctic meteorites. Lunar and Planetary Sci. XI, 1980, pp. 669-670.

Ming, D.W.: Chemical and Crystalline Properties of Minerals in the $MgO-CO_2-H_2O$ System. M.S. Thesis, Colorado State University, 1981.

Ming, D.W.; and Franklin, W.T.: Synthesis and characterization of lansfordite and nesquehonite. Soil Sci. Soc. Amer. Jour., vol. 49, 1985, pp. 1303-1308.

Nishiizumi, K.: Terrestrial and exposure histories of Antarctic meteorites. In International Workshop on Antarctic Meteorites, J.O. Annexstad, L. Schultz, and H. Wanke, eds., Lunar and Planetary Institute Tech. Report 86-01, 1986, pp. 71-73.

Schultz, L.: Terrestrial ages of Antarctic meteorites: Implications for concentration mechanisms. In International Workshop on Antarctic Meteorites, J.O. Annexstad, L. Schultz, and H. Wanke, eds., Lunar and Planetary Institute Tech. Report 86-01, 1986a, pp. 80-82.

Schultz, L.: Allende in Antarctica: Temperatures in Antarctic meteorites. Meteoritics, vol. 21, 1986b, p. 505.

Yabuki, H.; Okada, A.; and Shima, M.: Nesquehonite found on the Yamato 74371 meteorite. Sci. Papers Inst. Phys. Chem. Res., vol. 70, 1976, pp. 22-29.

MODELING THE GROWTH DYNAMICS OF FOUR CANDIDATE CROPS FOR CONTROLLED ECOLOGICAL LIFE SUPPORT SYSTEMS (CELSS)

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Tyler Volk, Ph.D
Academic Rank:	Assistant Professor
University & Department:	Department of Applied Science 26 Stuyvesant St. New York University N.Y. N.Y. 10003
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
Branch:	Advanced Research Projects Office
JSC Colleague:	Hatice Cullingford, Ph.D
Date:	August 14, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

The production of food for human life support for advanced space missions, such as the lunar and Mars bases, will require the management of many different crops. The research to design these food production capabilities along with the waste management to recycle human metabolic wastes and inedible plant components are parts of the NASA program in Controlled Ecological Life Support Systems (CELSS). Since complete operating CELSS have not yet been built, a useful adjunct to the research developing the various pieces of a CELSS are system simulation models that can examine what we currently know about the possible assembly of subsystems into a full CELSS. This report examines the growth dynamics of four crops--wheat, soybeans, potatoes, and lettuce--for their general similarities and differences within the context some of their important effects upon the dynamics of the gases, liquids, and solids in the CELSS.

Data for the four crops currently under active research in the CELSS program using high-production hydroponics are shown. Two differential equations, one each for the inedible and edible protions of the crop's biomass, are developed and applied to the general characteristics of each crop's growth pattern. Model parameters, such as ultimate sizes and growth rates, are determined by closely approximating each crop's data. These parameters are constant here for each crop, in order to demonstrate the capability for a relatively simple generic model to reproduce the overall characteristics of growth of different crops that can serve as a basis for including these crops in a model of a complete CELSS. In actuality these parameters are functions of environmental qualities, such as photosynthetic photon flux, photoperiod, atmospheric pCO_2 -- therefore further development along the lines based upon principles of photosynthesis and plant physiology is indicated. Models such as these can aid the engineering conceptual design of CELSS by providing flux rates of substances going into and leaving the plants. Flux rates for CO_2, H_2O, HNO_3, and O_2 are shown for the models developed here.

INTRODUCTION

The NASA CELSS program is developing Controlled Ecological Life Support Systems for advanced space missions involving long duration stays by humans (see CELSS, 1986). Simulation models help in the conceptual and preliminary engineering design of such systems by assembling the components as presently understood into a mathematical framework for asking and answering particular

Previous work along these lines by this author and John Rummel considered a CELSS that grows wheat as the sole crop (Volk and Rummel, Rummel and Volk, 1987). The model in these works used stoichiometries for various substances, such as plant protein and human urine, to develop balances to trace the flow of carbon, hydrogen, oxygen, and nitrigen through the various pathways. The model can grow wheat in a variety of planting schemes, between the end-points of a single batch (planting and harvesting once every fifty-five days) to much smaller plantings every day (yielding the same integrated amount of food production). Different planting schemes create different magnitudes of fluctuations in the standing biomass and in the buffer reservoirs of CO_2, H_2O, HNO_3, and O_2.

Eventually we need to extend such modeling efforts to crops other than wheat, to mimic what--due to human nutritional requirements--must be a mutliple-crop system. Several crops are undergoing tests in hydroponic systems to determine their growth characteristics in high-production systems with relatively high photosynthetic photon fluxes (PPF), atmospheric pCO_2 levels, optimized nutrient supply, controlled temperature and humidity, etc. Wheat is one of these crops; others include potatoes, soybeans, and lettuce. Questions regarding the differences in growth characteristics would affect the design of CELSS, such as how different will the growth environments have to be for different crops? This would affect the hardware design. Will each crop require a highly-customized system? Also of interest and a focus of this study is the comparative growth dynamics of the crops. How similar or how different are they? What do these similarities and differences mean for formulating the crops into a model of a CELSS?

FINDINGS

The study began with development of a questionnaire each crop researcher was asked to complete to provide a common set of characteristics for each crop. Answers were received both over the mail and by phone for each of the four crops requested: wheat, soybeans, potatoes, and lettuce. Examples of the information gathered included environmental conditions under particular high-yield experiments for both the aerial and root parts of the plants, and plant growth through time, food-type composition, planting and harvesting procedures, etc. Interested readers can contact me for more information.

Key data for this report are the growth curves of the four crops. Representative curves are shown in figures 1-5. Figure 1 shows typical soybean growth for the edible seed mass and inedible plant parts (inedible with respect to people), which in this case includes the leaves, stems, and roots. A division into edible and inedible biomasses is of fundamental importance in a CELLS because of the separation of material flow that would occur because of this division. The growth curves for edible and inedible portions of wheat (again the seeds as edible vs. all other plant

parts as inedible) is shown in figures 2 and 3, with figure 2 showing the growth of dry biomass for different light levels and figure 3 the growth of the fresh (wet) biomass and cumulative transpiration. Figure 4 shows the growth of tubers (the edible biomass) and the stems and leaves (the inedible biomass) for potatoes, both fresh and dry masses and for two diffferent photoperiods. Figure 5 shows several growth curves for lettuce.

The intention here is not to evaluate in any way the suitability for these crops for future CELSS. In fact, the assumption is that all these crops--indeed many more--would need to be grown for physically- and psycologically-satisfactory support of life. I have not attempted to provide all the environmental data for each crop; see me for more details. Note that some growth data is in grams per square meter and some in grams per plant. The goal at hand is to begin looking at these crops as a system, and what effects their various and special characteristics may have on the dynamics of a CELSS.

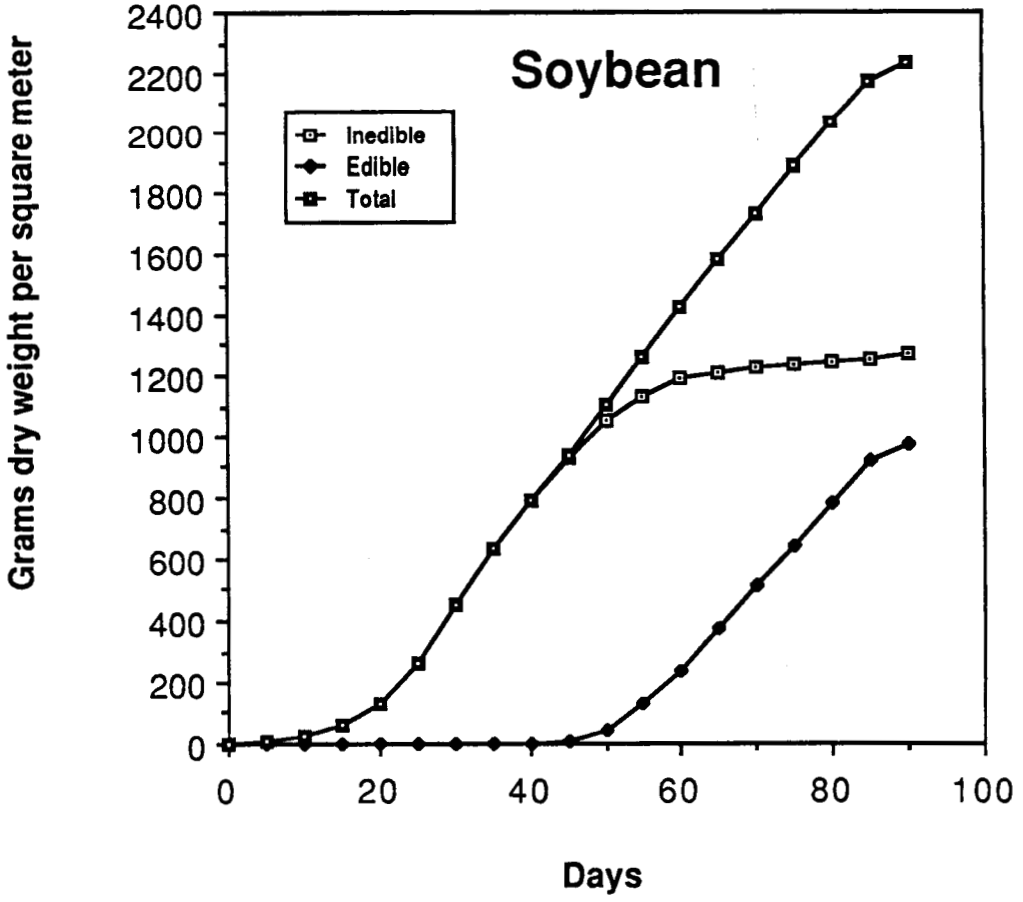


Figure 1. Soybean growth from data provided by D. Raper.

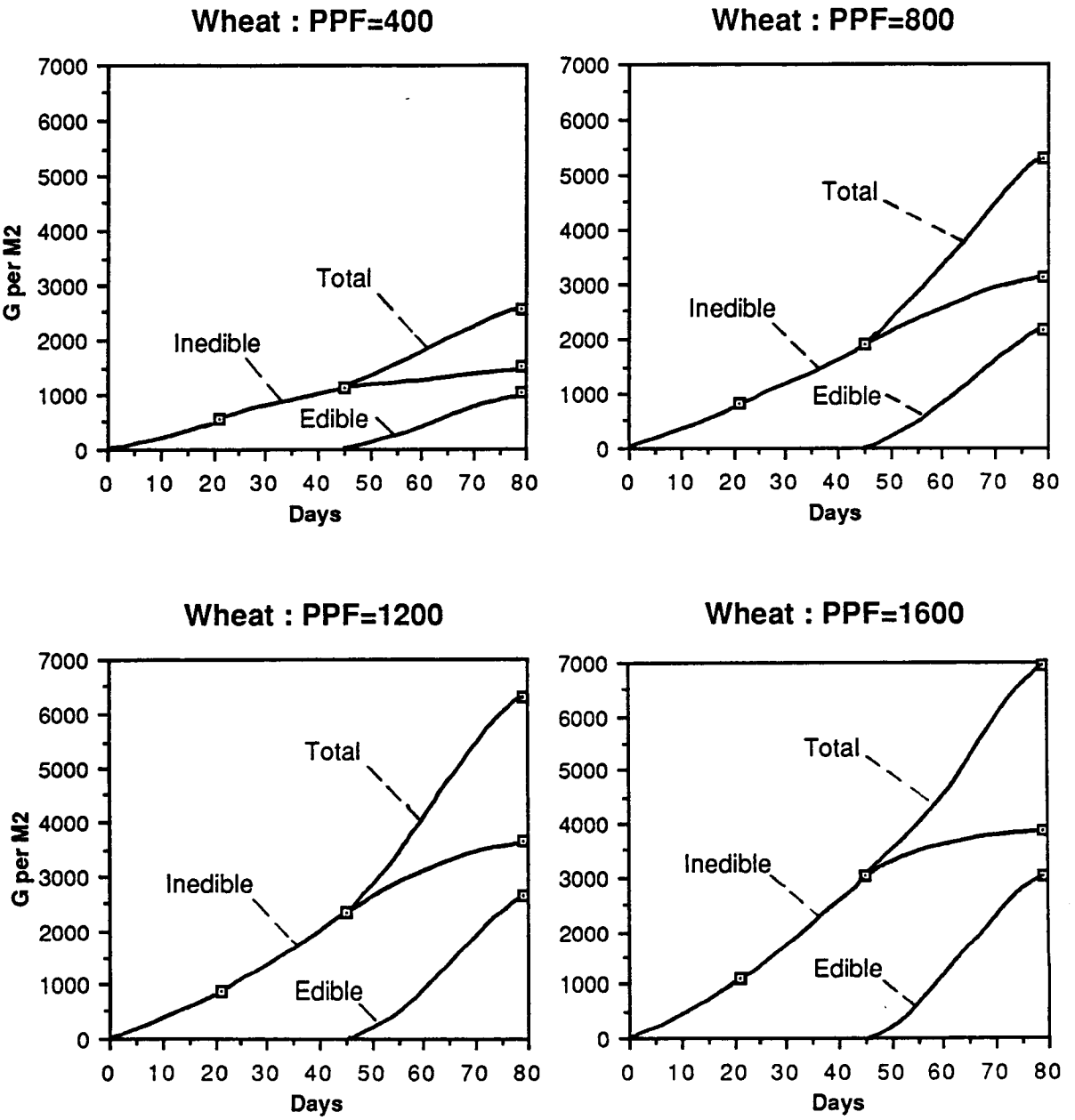


Figure 2. Wheat growth from data provided by B. Bugbee. See also Bugbee and Salisbury (1987). Some data points were modified by T. Volk in collaboration with B. Bugbee. Curves drawn by T. Volk. The photosynthetic photon flux (PPF) is in $\mu mol/m^2$-s.

C-4

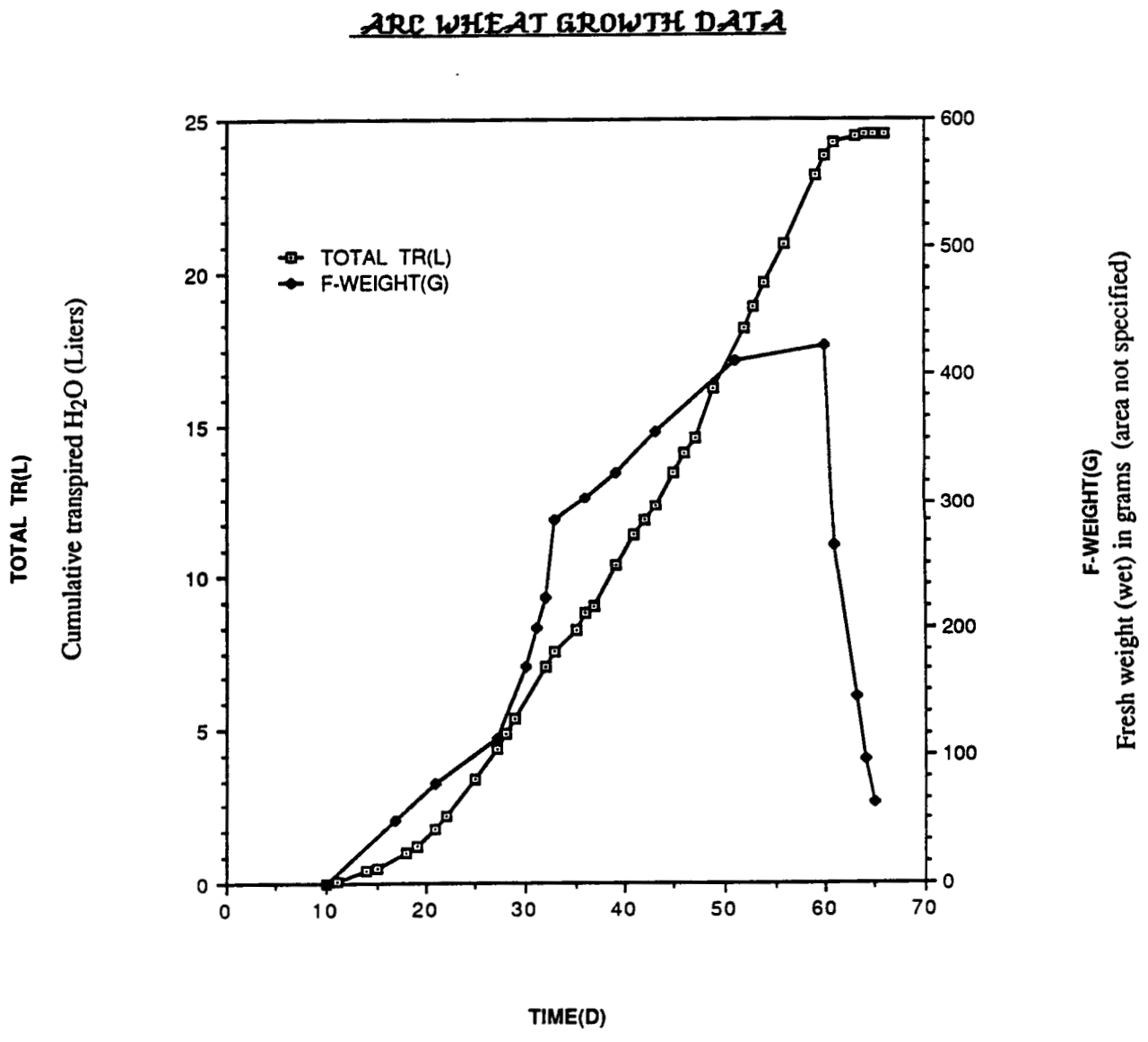


Figure 3. Wheat growth data from S. Schwartzkopf for cumulative transpiration and fresh-weight biomass.

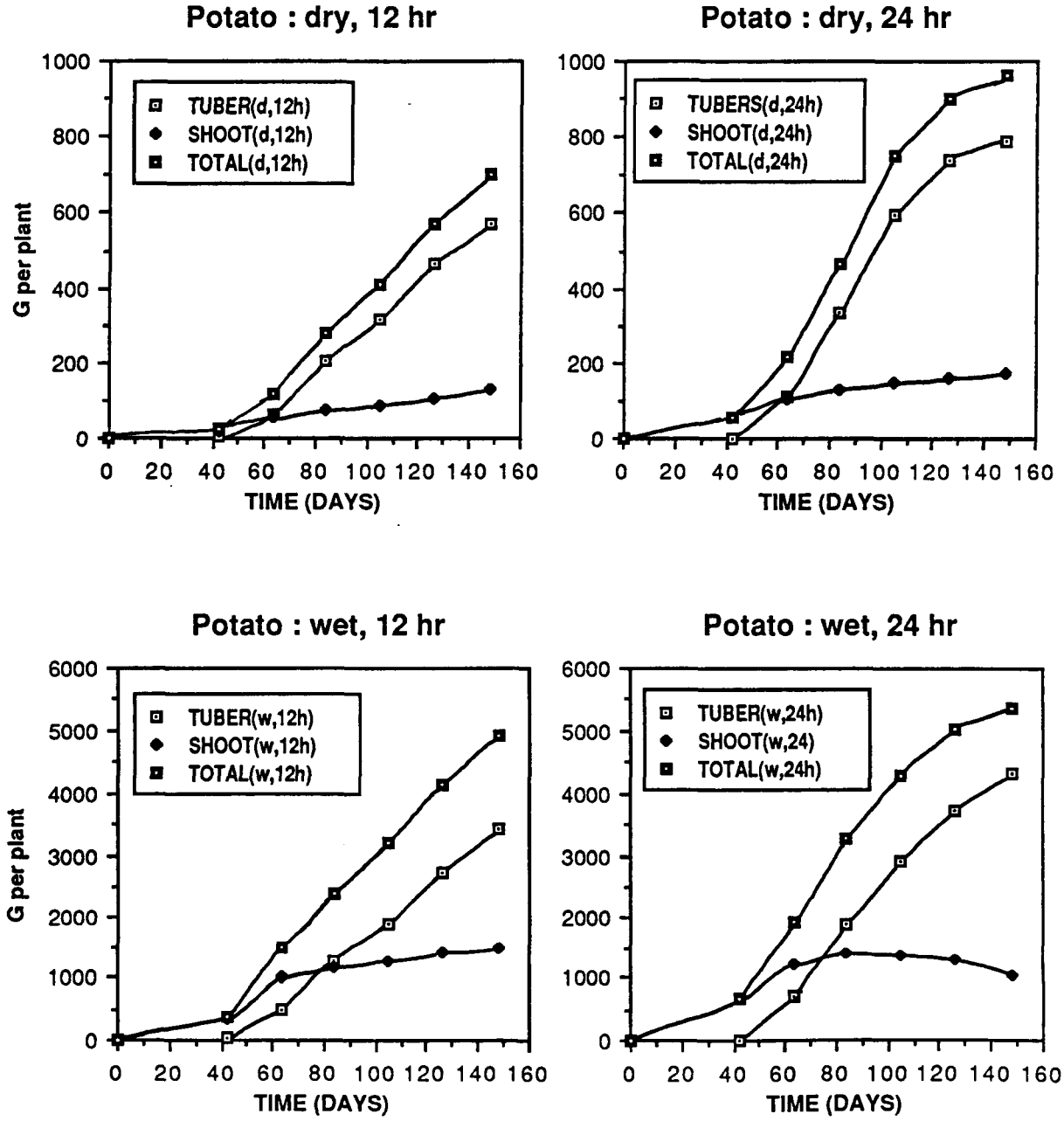


Figure 4. Potato growth from Wheeler and Tibbits (1987) for dry and fresh weights and for 12-hour and 24-hour photoperiods. The tuber curve is the edible mass and the shoot curve is the inedible mass.

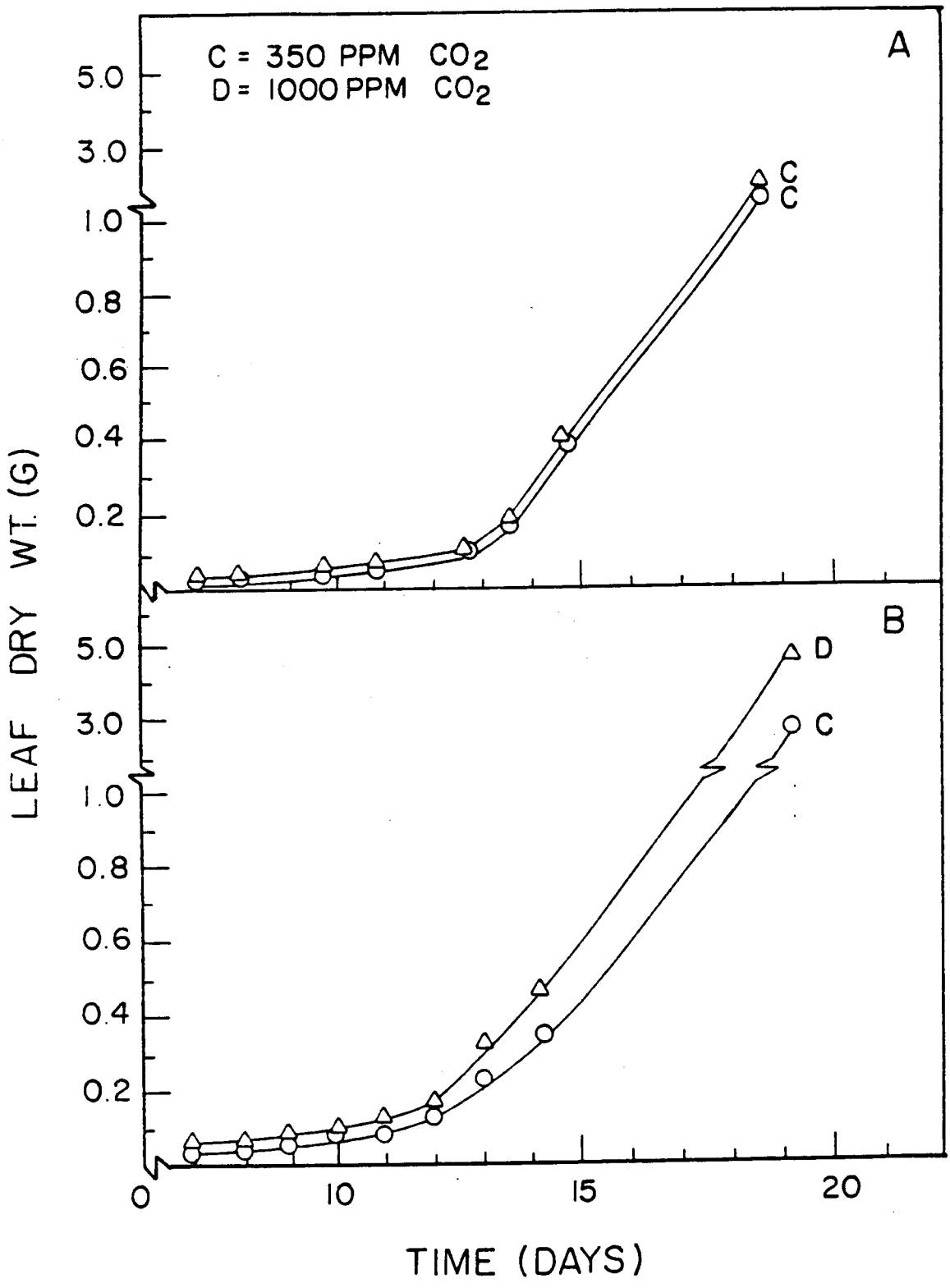


Figure 5. Growth profile of 'Waldmann's Green' leaf lettuce at 450 µmol/m^2-s of PAR + 350 µl/l CO_2 in 2 separate Minitron II chambers (top), and at 1000 µl/l CO_2 in one chamber and at 350 in another, both at the same PPFD (bottom). [Figure and figure description from Mitchell et al, 1986]

Many of the growth curves prominantly show the S-shaped or sigmoidal curve typical of biological systems. The logistic differential equation's solution imitates this S-shape of exponential growth followed by a leveling-off. In the logistic equation, $dC/dt = rC(1-C/K)$, where C is the biomass and t is time, two parameters appear: r and K. The r is the growth rate for the purely exponential part of the system. K, the carrying-capacity in a ecological system, in this case is the maximum biomass reached by the crop. I think of it as a "negative feedback from the lifetime", a environmentally-modifiable but inherently genetically-based slowing of the total growth rate (the dC/dt) by the approach of the crop to its mature size. The logistic equation, while about as simple as one could conceive to derive the S-curve, contains some biologically-meaningful parameters. This equation will be used for the growth of the inedible plant parts.

The equation for the edible plant parts must be somewhat differently structured. Like the inedible cells, the edible cells reproduce and total edible growth must contain a proportionality to the edible mass. But since the edible parts (except for lettuce--see below) are not producing their growing mass through photosynthesis, but rather receive products from photosynthesis of the inedible parts (the leaf mass), one would require the inedible biomass to also appear in the edible equation. Furthermore, as evident from the data in figures 1-5, the edible may begin substantially after the beginning of the inedible growth, and so a turning-on time (t*) is placed into the edible equation. In addition, since before t*, the edible mass is assumed equal to zero, an initial growth-spurt of edible is provided to ensure proper behavior (the term E_{min}). For further discussion of these equations, contact me

$$\frac{dM_{ined}}{dt} = r_{ined} M_{ined}\left(1 - \frac{M_{ined}}{K_{ined}}\right) \tag{1a}$$

$$t < t^*: \quad \frac{dM_{ed}}{dt} = 0 \tag{1b}$$

$$t > t^*: \quad \frac{dM_{ed}}{dt} = r_{ed} M_{ined}\left(\frac{E_{min} + M_{ed}}{K_{ed}}\right)\left(1 - \frac{M_{ed}}{K_{ed}}\right) \tag{1c}$$

The parameters t and t* are in units of time; r_{ined} and r_{ed} are in units of time^{-1}, and all other parameters are in consistent mass units. The system of eqns (1a-c) above was used for wheat, soybean, and potato. For lettuce the system was modified to be (with t* having a different meaning):

$$t < t^*: \frac{dM_{ed}}{dt} = r_{ed,1} M_{ed}\left(1 - \frac{M_{ed}}{K_{ed}}\right) \tag{2a}$$

$$t > t^*: \frac{dM_{ed}}{dt} = r_{ed,2} M_{ed}\left(1 - \frac{M_{ed}}{K_{ed}}\right) \quad (2b)$$

$$\frac{dM_{ined}}{dt} = \frac{dM_{ed}}{dt} \frac{K_{ined}}{K_{ed}} \quad (2c)$$

These systems of equations (1a-c) and (2a-c) were placed into the STELLA modeling program for the MacIntosh computer (see figure 6). STELLA is useful for relatively simple dynamic models for the user creates a digram in the program that corresponds to the model equations, and a hierarchical structure allows one to "open-up", examine, and change the various components at will. The graphic quality allows the user to recall the model structure after not using the model for a period of time, facilitating a very flexible interaction. I envision such modeling systems as STELLA to be very useful to models at the stage of CELSS when large-scale design differences need to be explored and analyzed by a number of different types of people.

The program was run and graphic output of edible and inedible biomass through time generated with the same scales as the crop data. I shifted parameters in ways that made sense to me until the model generated approximately what the data showed. No attempt was made to optimize the fit--there is little to be gained by this at this point. The parameters used for each crop is listed in Table 1 and the model outputs shown in figure 7.

Note that the crop data imitated by the figure 7 models for wheat was the figure 2 (PPF=1200) case and for potatoes was the figure 4 (dry, 24 hour) case.

The model curves demonstrate that it is relatively easy to imitate the data with a single model whose parameters have some biological meaning in at least a crude sense. Table 1 lists the actual planting masses for the crops, but I need to investigate whether the data at t=0 actual means the initiation of the crop from seed or tissue, or is the transplanting time after initial seeding growth. Some further adjustment to account for additional information on the meaning of time t=0 may therefore be necessary.

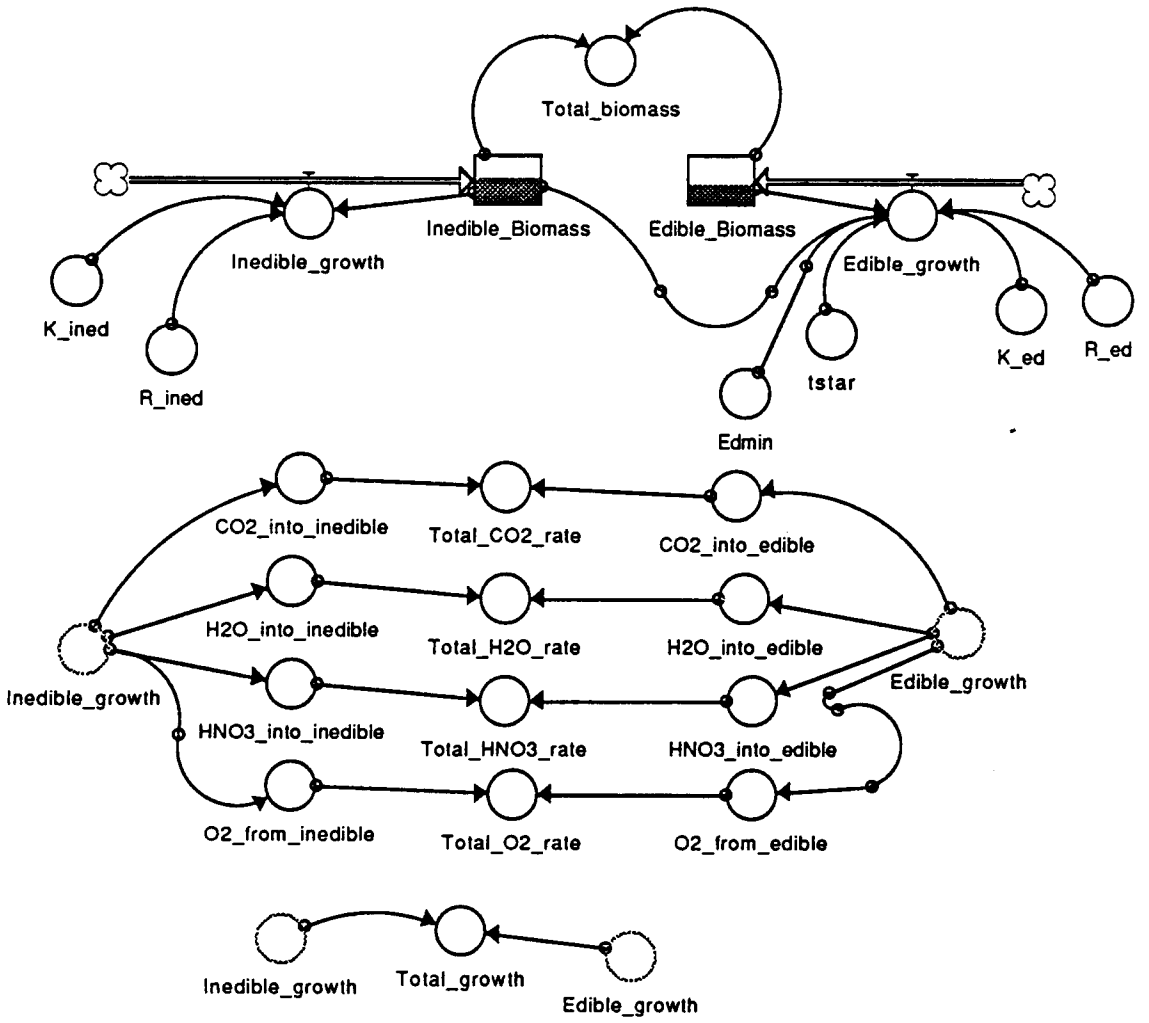


Figure 6. Crop model structure in STELLA programming language. Upper portion of diagram corresponds to eqns (1a-c).

Table 1. Parameters for Crop Models

Parameter	Wheat	Soybean	Potato	Lettuce
r_{ined} (day^{-1})	0.09	0.10	0.06	0.2 r_{ed}
r_{ed} (day^{-1})	0.15	0.10	0.20	0.2 to 0.5
K_{ined}	3700.0	1300.0	200.0	2.0
K_{ed}	3000.0	1100.0	800.0	20.0
E_{min}	80.0	80.0	80.0	X
$M_{ined,o}$	150.0	20.0	5.0	X
$M_{ed,o}$	0.0	0.0	0.0	0.008
t* (days)	45.0	45.0	40.0	11.0 (r_{ed} switch)
[actual $M_{ined,o}$ in seeds or tissue]	70.0	2.0	0.15*	0.008**

* not certain whether wet or dry
**not certain whether 0.008 or 0.0008

Note different units between crops for K_{ined} ; K_{ed} ; K_{ined} ; $M_{ined,o}$; $M_{ed,o}$
For wheat and soybean all are in grams dry mass per square meter.
For potato and lettuce all are in grams dry mass per plant (approx. 5 plants per square meter for potato).

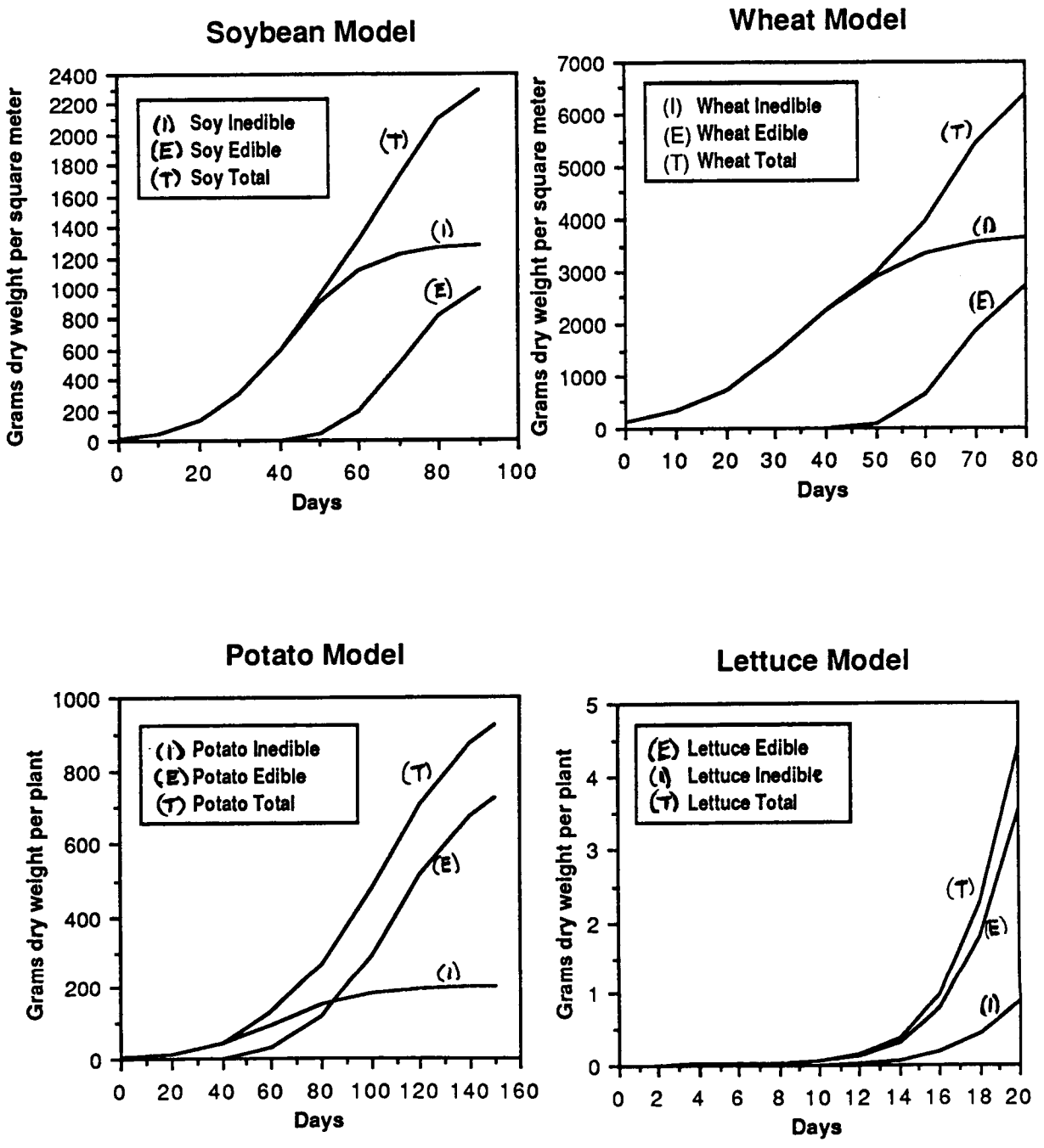


Figure 7. Models for the four crops. Eqns (1a-c) were used for soybean, wheat, and potato models. Eqns (2a-c) were used for lettuce. Parameters are in Table 2. (Curves are smoother than shown here.) The wheat parameters were adjusted to obtain curves close to those of the data with "PPF=1200" in Figure 2. The potato parameters were adjusted to obtain curves close to those of the data with "dry, 24 hr" in Figure 4. Closer fits are obviously possible; but the purpose at hand is to demonstrate the potentials of a generic crop model.

Volk and Rummel (1987) list formulas for protein, carbohydrate, lipid, fiber, and lignin that can be placed into balanced equations for carbon, hydrogen, oxygen, and nitrogen. It is therefore possible to calculate the uptake of CO_2, H_2O, and HNO_3, and the production of O_2 by the crops. Per unit mass of biomass, these compounds vary as a function of the biomass's fractional distribution of protein, carbohydrate, lipid, fiber, and lignin. The details of the calculation is not shown here; please contact me if interested. Table 2 shows the mass balances for the four crop models. Note the substantial differences betweeen, for example, the CO_2 required and O_2 produced per gram of edible soybean vs. per gram of edible wheat. This difference is due primarily to the difference in lipid content. Such differences represent differences in the fluxes of these materials between the crops and their environments, and will presumably be important in the engineered hardware designs with respect to how similar or different the hardware must be for the various crops.

The balances in table 2 were used with the crop growth models (see program diagram in figure 6) to calculate the fluxes of CO_2, H_2O, HNO_3, and O_2 during growth; these fluxes are shown in figure 8. Note the different shapes for the crops. Such actual curves will be known during the operation of a CELSS (for example, since CO_2 will be monitored and maintained at desired levels in the crop's atmosphere, the amount of CO_2 injected to maintain these levels will be known). Due to the characteristic patterns of these fluxes, it may be possible to tie this knowledge into monitoring systems of state of the whole crop. Note that these curves assume a constant percentage of protein, carbohydrate, lipid, fiber, and lignin for the edible and inedible during their respective growths. This is clearly not the case as seen in the decrease in leaf nitrogen in the hydroponic wheat during seed growth (Bugbee and Salisbury, 1987). Obviously the next step is to let this nitrogen difference represent a decrease in the edible parts' protein in the late stage of growth, and it would be informative to see how much this affects the CO_2, H_2O, HNO_3, and O_2 fluxes. We could tell how much uncertainty in composition over time affects the fluxes, and therefore if we know how accurately the fluxes need to be predicted, this would give requirements for the data.

Table 2. Mass Balances for Crop Models

Mass Types		Wheat	Soybean	Potato	Lettuce
Edible Mass Fractions					
	protein	0.21	0.45	0.13	0.26
	digest. carbo.	0.74	0.30	0.84	0.12
	lipid	0.02	0.25	0.00	0.06
	fiber	0.03	*	0.03	0.56
	lignin	0.00	*	0.00	0.00
Edible Mass Balances (g/g-dry-biomass)					
	CO_2 (in)	1.651	2.102	1.572	1.822
	H_2O (in)	0.582	0.662	0.585	0.570
	HNO_3 (in)	0.160	0.343	0.099	0.198
	O_2 (out)	1.393	2.107	1.256	1.590
Inedible Mass Fractions					
	protein	0.09	0.17	0.19	0.11**
	digest. carbo.	0.14	0.80	0.30	0.11**
	lipid	0.00	0.03	0.00	0.00**
	fiber	0.72	*	0.45	0.78**
	lignin	0.05	*	0.06	0.00**
Inedible Mass Balances (g/g-dry-biomass)					
	CO_2 (in)	1.720	1.632	1.755	1.681
	H_2O (in)	0.561	0.595	0.556	0.554
	HNO_3 (in)	0.068	0.129	0.144	0.084
	O_2 (out)	1.349	1.356	1.455	1.319

* fiber and lignin were included in the soybean carbohydrate data
** values were assumed by T. Volk

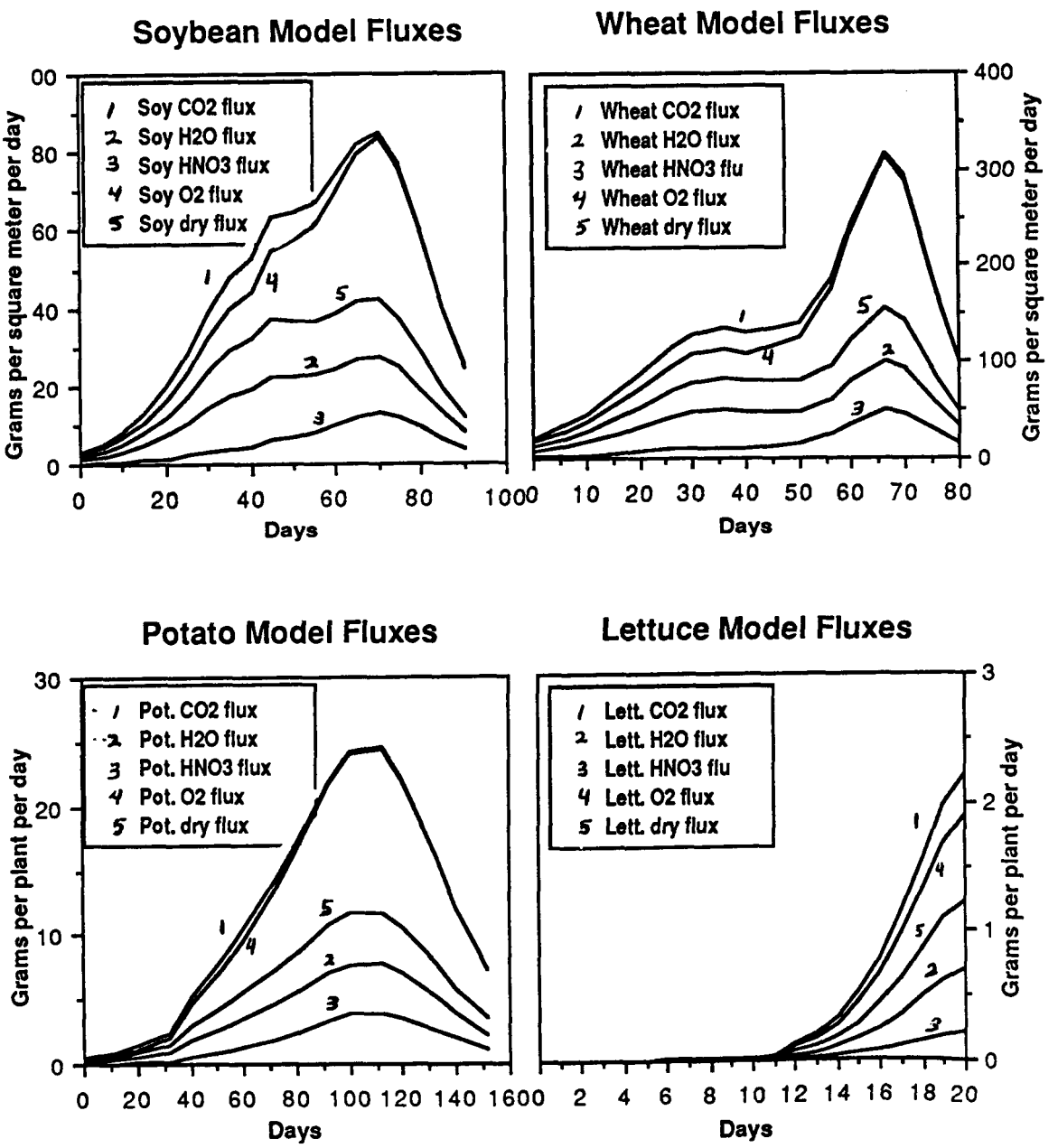


Figure 8. Fluxes for the four crop models: CO_2, metabolic H_2O, nutrient HNO_3, O_2 produced, and total dry weight biomass (edible plus inedible). Note different units for the different crops. Fluxes are from the models of Figure 7 using the stoichiometries of Table 2.

CONCLUSION

A simple generic model has been demonstrated that can emulate the growth dynamics of four different candidate crops for Controlled Ecological Life Support Systems and provide mass fluxes associated with these growing crops for incorporation into a whole-system CELSS model . That the model at this stage is simple is significant; an initial simplicity is desirable because the model will tend to quickly become more complex when it incorporates additional refinements, particularly sensitivities to environmental variables. There is every reason to expect that a generic model like the one demonstrated here will be useful in constructing a model system for studying the dynamics of a space farm. Combining the crops into such a farm could be a subsequent step from this study.

Another next step is to incorporate the above-mentioned refinements. It is obvious that by adjusting the parameters in eqns (1a-c), fits to the other growth curves shown for wheat and potatoes in Figures 2 and 3 could be obtained. The model parameters, such as growth rates, r_i's, and ultimate biomasses, K_i's, are not constant, but must be functions of environmental conditions. A reasonable approach would be to develop the environmental functionality of these parameters along lines of classical mathematical treatments of photosynthesis, such as in Gates (1980), wherever possible. That way the data--for example, the variation of growth with light shown in figures 2 and 4 for wheat and potatoes--would not be used for fitting, but rather for model validation. Transpiration submodels and the relationship between atmospheric pCO2, humidity, nutrient uptake, and growth, need to be developed in order for the various design tradeoffs between energy, volume, etc. to be investigated. The model shown here could serve as a basis for further development.

REFERENCES

Bugbee, B. and F. B. Salisbury (1987) Exploring the limits of crop productivity: photosynthetic and carbon partitioning efficiency in an optimizing environment (draft of manuscript in preparation).

CELSS (1986) Controlled Ecological Life Support Systems: CELSS '85 Workshop, edited by R. D. MacElroy, N. V. Martello, and D. T. Smernoff, NASA-TM-88215

Gates, D. M. (1980) Biophysical Ecology, Springer-Verlag, New York.

Mitchell, C. A., S. L. Knight, and T. L. Ford (1986) Optimization of controlled environments for hydroponic production of leaf lettuce for human life support in space, in Controlled Ecological Life Support Systems: CELSS '85 Workshop, edited by R. D. MacElroy, N. V. Martello, and D. T. Smernoff, NASA-TM-88215, pp. 499-521.

Rummel, J. D. and T. Volk (1987) A modular BLSS simulation model, Advances in Space Research, in press.

Volk, T. and J. D. Rummel (1987) Mass balances for a biological life support simulation model, Advances in Space Research, in press.

Wheeler, R. M. and T. W. Tibbits (1987) Utilization of potatoes for life support systems in space: III. Productivity at successive harvest dates under 12-h and 24-h photoperiods, American Potato Journal, 64, pp. 311-320.

RAMAN SPECTRA OF ADSORBED LAYERS ON SPACE SHUTTLE AND AOTV THERMAL PROTECTION SYSTEM SURFACE

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Ronald J. Willey, Ph.D.
Academic Rank	Assistant Professor
University & Department	Northeastern University Department of Chem. Eng. Boston, MA 02115
NASA/JSC	
Directorate:	Engineering
Division:	Structures & Mechanics
Branch:	Thermal
JSC Colleagues:	John E. Grimaud Carl D. Scott, Ph.D.
Date:	3 September 1987
Contact Number:	NGT 44-001-800

ABSTRACT

Surfaces of interest to space vehicle heat shield design were struck by a 2 W argon ion laser line while subjected to supersonic arc jet flow conditions. Emission spectra were taken at 90^{o} to the angle of laser incidence on the test object. Results showed possible weak Raman shifts which could not be directly tied to any particular parameter such as surface temperature (range 900 to 1,500 K), or gas compositions (air, $O_2./N_2$ mixture, N_2, and argon).

The investigation must be considered exploratory in terms of findings. Many undesirable effects were found and corrected as the project progressed. For instance, initial spectra settings lead to ghosts (stray light within the spectrometer) which were eliminated by closing the intermediate of filter slit of the Spex from 8 mm to 3 mm. Further, under certain conditions, plasma lines from the laser were observed. Several materials were also investigated at room temperature for Raman shifts. Results showed Raman shifts for RCC and TEOS coated materials. HRSI materials showed only weak Raman shifts, however, substantial efforts were made in studying these materials. Baseline materials showed the technique to be sound (Raman spectra of quartz and a silicone grease were checked). The original goal was to find a Raman shift for the HRSI RCG coated material and tie the amplitude of this peak to Arc jet conditions. Weak Raman shifts may be present, however, time limitations preventive confirmation.

BACKGROUND

The Raman effect involves inelastic collisions of a beam of light (photons) which strike a surface or gas molecule. The photons, which strikes the surface, can give up of a portion of their energy to the surface molecules (in a quantum amount which is characteristic of the surface molecules). These photons then exit the surface at lower energy and their frequency is "shifted" to a lower wavenumber. The Raman effect or Raman shift was first predicted by Adolf Smekal in 1923. Sir Chandrasekhara Venkata Raman experimentally demonstrated the shift around 1927. The effect is quite weak with an intensity of about 1/100,000 of the incident beam (most collisions are elastic).

The application of Raman spectroscopy to surface science has expanded tremendously in the past 5 years. Improvements in data acquisition by microcomputers and in high speed A/D converters have opened up the ability to record low power light sources such as those generated by a Raman shift. Surface Raman spectroscopic techniques have been applied to many materials primarily of catalytic nature. Recent works include the adsorption of methane onto silica (Morrow, 1980) and the characterization of nickel oxide states on alumina catalysts (Wachs et al., 1986).

Raman spectroscopy ties into the Johnson Space Center Atmospheric Reentry Materials and Structures Evaluation Facility (Arc Jet) because of recent acquisition of a 5 W argon ion laser and a Spex Model 1877 triple spectrometer. With this equipment, surface Raman is a possible technique to use to characterize and better understand reusable heatshield materials under conditions which simulate reentry.

The design of reusable heatshield materials require 3 major design criteria: stability, ability to shed radiation to the surroundings, and be non-catalytic. It is the third criteria, a non-catalytic surface, where surface Raman may provide fundamental information. The current workhorse heatshield material is high temperature reusable surface insulation (HRSI) which is composed of alumina and silica fibers coated with a reaction cured borosilicate glass (RCG). The RCG coating is an excellent radiator and is pretty much non-catalytic, however, a small amount of atom recombination can occur (Kolodziej and Stewart, 1987). Therefore, heat shield design must take into account this factor. The problem however is that fundamental mechanisms have been lacking.

Recently, Seward, 1985, has proposed a recombination mechanism which predicts an interesting and critical result: a peak in the recombination rate with temperature. Seward's model is based on the following mechanism:

$$Si + Og \leftrightarrows Si{=}O \qquad (1)$$

$$Si{=}O + Og \longrightarrow Si + O_2 \qquad (2)$$

and his resultant model:

$$\text{recombination coefficient} = \frac{2\ P\ So\ N \exp(-E/kT)}{So\ N + P\ N \exp(-E/kT) + Delta}$$

where:

So is the sticking coefficient

N is the atomic flux to the surface

P is the steric factor

k is the Boltzmann constant

T is the temperature

E is the activation energy for the forward reaction 2

Delta is thermal desorption rate which is a function of temperature = Ca (kT/h) exp (-D/k T)

where:

Ca is the number of adsorption sites per unit area

D is the thermal desorption energy

Equation two predicts a transition from zero order to first order for the recombination coefficient in oxygen partial pressure when the thermal desorption begins to be important (the reverse Reaction 1). It should be noted that N depends on the first order of oxygen concentration. Further, because of the predicted Arrhenius behavior for thermal desorption, the predicted recombination coefficient takes a nose dives to practically zero above its peak temperature. Experimentally, Kolodziej and Stewart, 1987, have reported recombination coefficients dropping above a certain temperature. Thus an initial objective of this work was to see if a Raman shift could be found which was associated with Si-O bonds on HRSI materials, and further, how did the Si-O bond alter under various arc jet conditions such as current and gas compositions.

EXPERIMENTAL

EQUIPMENT DESCRIPTION

Figure 1 represents a schematic diagram of the equipment used in this work. At the heart of the work was Model Spex 1877 0.6 m triple spectrometer. Light from the arc tunnel was focused through a 38 OD 75 fl lens through a 400 nm low wavelength filter past a mirror through another lens (a 76 OD 250 fl) into the entrance slit of the spectrometer. A 150 groove/mm filter grating filtered light before it reached the main spectrometer entrance slit. For this work the main spectrometer grating was set at 600 grooves/mm (two other settings are possible). Light diffracted by the

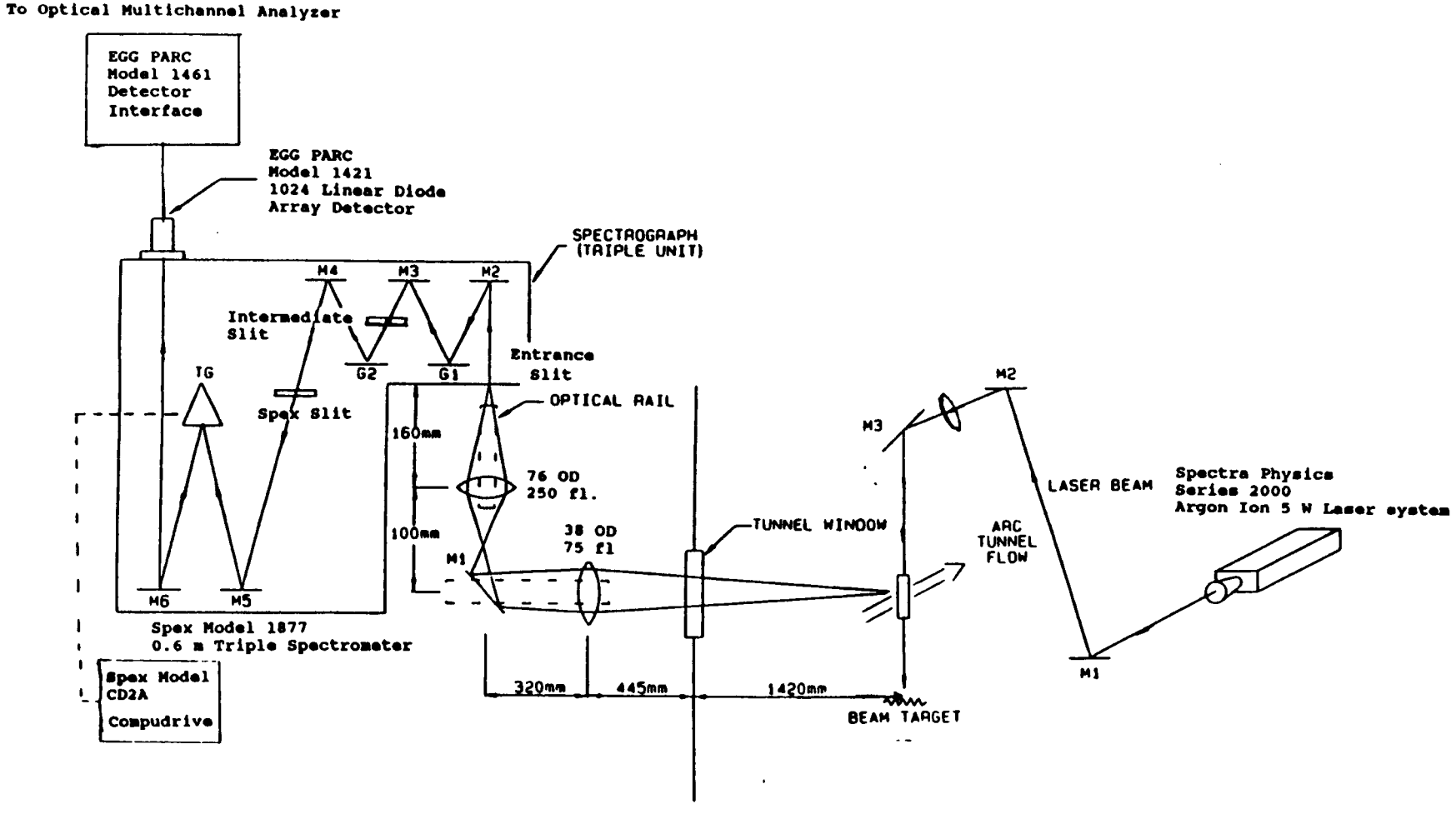


Figure 1. Apparatus Diagram.

spectrometer was collected by a EG&G PARC Model 1421 diode array which consisted of 1024 diodes sensitive in the UV and visible range. The diode array was interfaced to an optical multichannel analyzer (OMA) via a EG&G PARC Model 1461 detector interface. The OMA allowed the collection of spectroscopic data onto hard or soft disk for latter analysis such as plotting by a HP 7440 plotter or intensity analysis by a Epson FX 85 printer. The OMA also controlled the spectrograph wavelength setting (via a BASIC program written by Dr. Fred Wierum), scan times, and number of scans collected. Minimum scan time was 16.63 milliseconds across a wavelength range of 70 nm. Up to about 30,000 scans per collection period were possible.

Raman scattering was made possible by the use of a Spectra Physics Series 2000 argon ion laser system. The laser had two major excitation lines at 514.536 nm and 488.0 nm. The beam for the laser was focused into the centerline of the arcjet tunnel by 3 mirrors and one lens with a focal point of 1.5 m.

OPERATION

An example experimental start-up and operation procedure will be presented. The first step involved turning on the Spex spectrograph and entering the wave length into Spex CD2A compudrive to insure that the grating position matched the compudrive position. Then two configuration changes were made on the compudrive. Next, the N_2 purge gas flow was checked and set at 130 units on the rotameter. Then, the Model 1461 A/D converter was turned on. Then, one proceeded to the control room to turn on the printer, the plotter, and then the OMA system. The software was booted up and the system now ready. Full detailed directions have been written by Dr. Wierum and should be followed.

The laser control box was turned on lifting the main electrical switch up. Water flow was turned on and the water flow adjusted to 3 gallons per minute. After a 10 minute warm up the magnetic current was set at 9 amps and the current was set to about 38 amps. Under these conditions the laser would emit a line at 514.536 nm of about 2 watts in power. The laser line was then focused into the arc jet by adjusting each mirror in the optical train working from the laser source. Final focusing was done by adjusting the laser line intersection point with the optical path of the spectrometer and hitting the desired target (a space shuttle tile). Special argon laser line protective glasses were worn during the laser operation and mirror adjustment procedure.

For runs involving the arc jet, gas flow rates were set at 0.045 kg/sec (0.1 lb/sec). Currents ranged from 400 to 1200 amps depending on what conditions were being evaluated. The heated gases flowed through a 2.25 inch diameter throat and exited through a 15 inch diameter nozzle set at a 15 degree half angle. Under the condition studied the exit velocity was about 4,000 m/s at a Mach number of 7.0. Test articles evaluated in the arc jet flow were a 4 1/2 " X 5 " wedge inserted into a copper holder and a 2" radius HRSI RCG coated hemisphere model. Run numbers for this work were 2-563-SD to SD-574-SD. Further run details may be obtained from arc jet personnel. Table 1 is a brief summary of arc jet conditions run.

The laser used was a Argon Ion laser with a tunable excitation lines of 488 or 514.5 nm. Lower powered laser lines are also available at 457.9, 465.8, 472.7, 476.5, 496.5 and 501.7 nm.

TABLE 1. SUMMARY OF ARC JET CONDITIONS EVALUATED FOR RAMAN SHIFTS ON HRSI

Test Articles:

HRSI Wedge 4 1/2" X 5" inserted into a copper leading edge holder

HRSI Hemisphere of 2" in radius

Arc Jet Conditions:

Mass Flow Rate kg/sec	Gas Composition Mass %			Current Amps	Power MW	Enthalpy MJ/kg	Typical Surface Temperatures, Kelvin	
	O_2	N_2	Ar				Wedge	Hemisphere
AIR								
0.0454	24.00	76.00	0.00	400.00	1.03	11.88		1429.44
0.0454	24.00	76.00	0.00	500.00	1.24	15.24	962.78	1477.78
0.0454	24.00	76.00	0.00	700.00	1.61	18.00	1002.22	
0.0454	24.00	76.00	0.00	900.00	1.98	18.62	1042.78	
0.0454	24.00	76.00	0.00	1200.00	2.52	19.03	1068.89	
Mixtures								
0.0454	6.00	94.00	0.00	500.00	1.31	17.23	932.78	
0.0454	10.00	90.00	0.00	500.00	1.30	17.53	914.44	
0.0454	21.60	68.40	10.00	500.00	1.26	11.50	941.67	
Nitrogen								
0.0454	0.00	100.00	0.00	400.00	1.08	11.42		1291.11
0.0454	0.00	100.00	0.00	500.00	1.32	15.91	956.11	1446.67
0.0454	0.00	100.00	0.00	900.00	2.12	22.46	1011.67	
0.0454	0.00	100.00	0.00	1200.00	2.52	19.03	1068.89	
Argon								
0.0454	0.00	0.00	100.00	500.00	0.12		797.78	553.89

Alignment was accomplished by focusing a helium-neon laser through the Spex optical path onto the test article located in the test position. This was made easy by using the Spex's side exit for the helium-neon laser and setting the Spex at the laser's wavelength (632.8 nm). Once the "red dot" was located, the "green" line from the argon laser was focused through its mirror system down into the test chamber. Final alignment required that the "green dot" strike the "red dot" on the test article. This approach assured that the Spex would be looking at a region excited by the argon laser line.

After alignment, the neon laser was turned off and the side entrance closed. The side exit mirror was then changed to straight pass. The spectrometer was then reset to a centering wavelength of 560 nm (280 on the compudrive and 140 on the filter setting). The entrance slit was adjusted to 1.0 mm, the intermediate slit to 2.8 mm and the Spex slit to 600 microns. For these settings with a 514.6 nm laser excitation line, the wavelength range is 240 to 2500 cm^{-1}. For a wavenumber range of 2,200 cm^{-1} to 4,000 cm^{-1} the Spex was reset to a center wavelength of 616 nm (308 on the compudrive and 154 on the filter). The intermediate slit was reset to 3.2 mm.

Several operating conditions for the OMA were examined. Scan exposure times of 16.63 ms to 12,000 ms were used. Longer exposure times eliminated noise, however, one must watch for overloading the photo diode array (>16,000 counts per scan). Multiple collections are also possible. A range of 1 to 30,000 scans could be collected. The best operating condition found was an exposure at 100 ms for 100 or 1,000 scans.

FINDINGS

CALIBRATION

Scale calibration was accomplished by calculating wavenumbers for neon lamp lines at various Spex settings. The formula used was:

Shift in m^{-1} = $(1/\lambda\ m^{-1}$ excitation) - $(1/\lambda\ m^{-1}$ neon)

$cm^{-1} = 100\ m^{-1}$

Because of the non linear nature of wavenumber to wavelength conversion, a cubic fit available on the OMA system was used. The summary of calibrations is presented in Table 2.

PRELIMINARY FINDINGS

After 2 months of preparation, the laser was turned on and the first set of Raman spectra were collected under arc jet conditions. A review of results showed narrow peaks of very high counts (2×10^6 at 16.63 seconds total exposure time). The top line in Figure 2 shows the narrow peaks observed. These peaks disappeared when the intermediate filter slit was closed from 8 mm to 7 mm (indicating stray light inside the Spex). Because of stray light, the Spex intermediate slit was closed down to 2.8 to 3.2 nm so that only the 70 nm required to fill the photo diode array would pass into the main spectrometer. A disadvantage to centering the filter to the Spex setting is

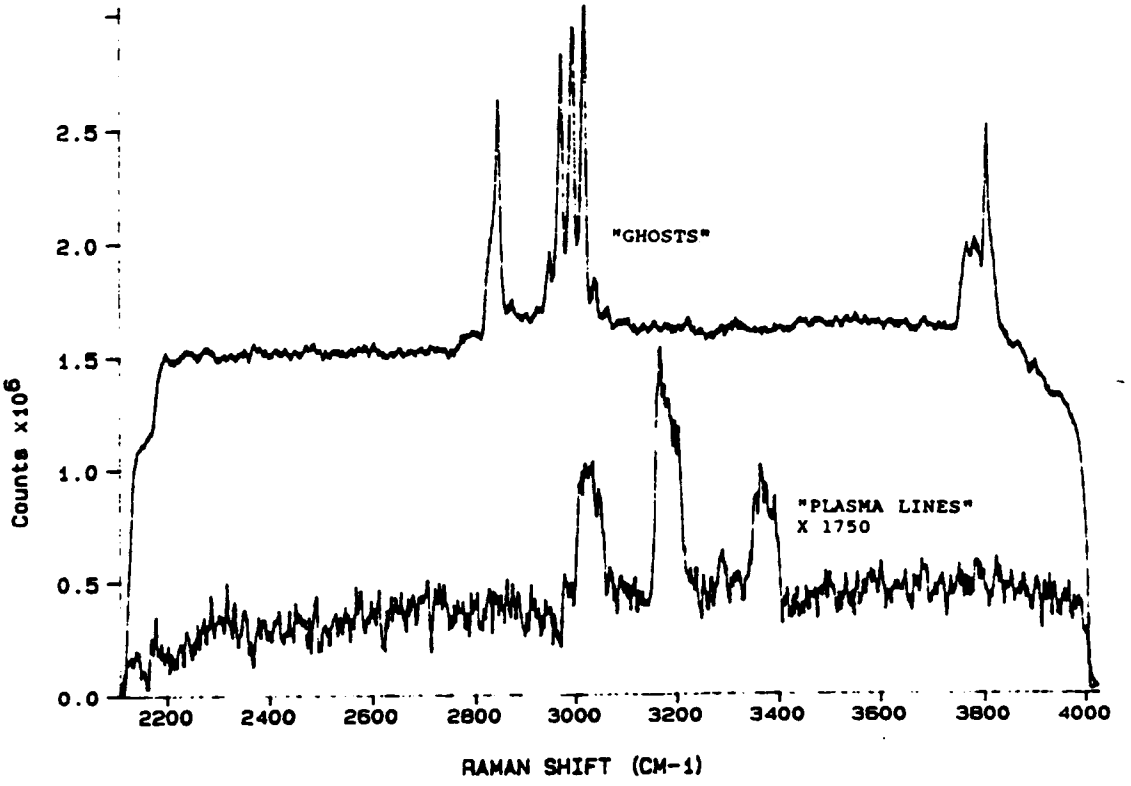


Figure 2. Ghosts and Plasma.

that it removes the ability to change wavelength remotely. Any grating change thus requires an arc off during a run.

Shown in the bottom half of Figure 2 is another anomaly discovered in the very weak spectrums - plasma lines. Plasma lines are atomic lines emitted from the laser whether it is lasing or not. These lines can be eliminated by a Model 1460 tunable excitation filter available from Spex Industries of Edison New Jersey. Time prevented the employment of a filter, however, it is recommended that one be purchased and put into use. To eliminate the influence of plasma lines in Raman spectra, background was taken in which the laser was left on however it was adjusted slightly off line so that it did not lase. Spectrum shown in the figures discussed below are for a background which includes the plasma radiation. Background spectra are referred to as plasma background.

TABLE 2. WAVENUMBER CALIBRATION DATA (Units cm^{-1})

Laser Excitation Line = 514.5 nm

OMA File # on Drive 1	Centering Wavelength nm	A0	A1	A2 ($X 10^4$)	A3 ($X 10^7$)
636	560	237.035	2.76972	-8.062	3.206
637	616	2106.73	2.17466	-4.304	1.349
639	672	3642.32	1.75655	-1.875	9.316E-3

Laser Excitation Line = 488.0 nm

638	527	98.7479	2.91402	-4.557	0.7710
671	560	1314.23	2.65549	-6.100	2.205

MATERIAL SURVEY

Further investigation indicated the technique had to be checked. Dow Corning Silicon Grease, DC-11, was used for this check. DC-11's infrared spectrum, completed by Mr. Dwayne Caston of Lockheed, is shown in Figure 3. Peaks are seen at wavenumbers in cm^{-1} of 670 (s), 800 (b), 870 (w), 1100 (b), 1255 (m), 1410 (s), 1440 (w), 1600 (w), 1940 (w), 2020 (w), 2500 (w), 2900 (s) and 2950 (s). The small letters after the number have the following meaning: (s) for a strong peak, (m) for a moderate peak, (w) for a weak peak, and (b) for a broad peak. A Raman spectrum in the same wave number range on DC-11 is shown in Figures 4 & 5 (Raman spectrums for other materials are also shown in these figures). Peaks are seen in the Raman spectrum at wavenumbers of 514(s) (this is below the infrared results), 711 (s), 806 (w), 886 (w), 1123 (w), 1234 (w), 1415 (m), 2916(s), 2979 (s) cm^{-1}. This is an excellent match, however, relative

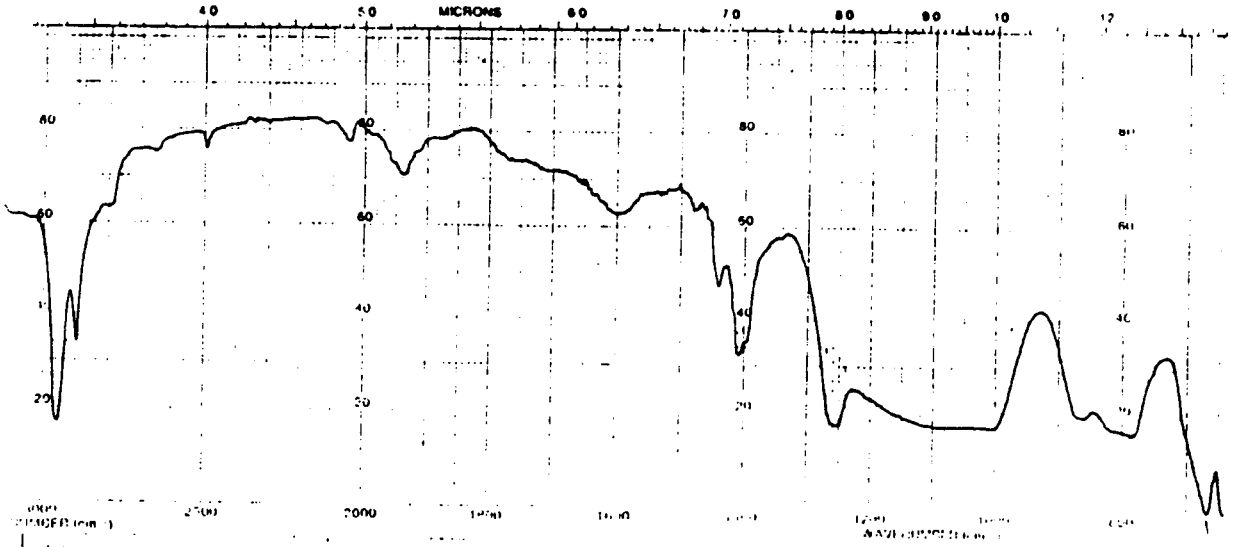


Figure 3. Infrared Spectrum of Dow Corning grease DC-11.

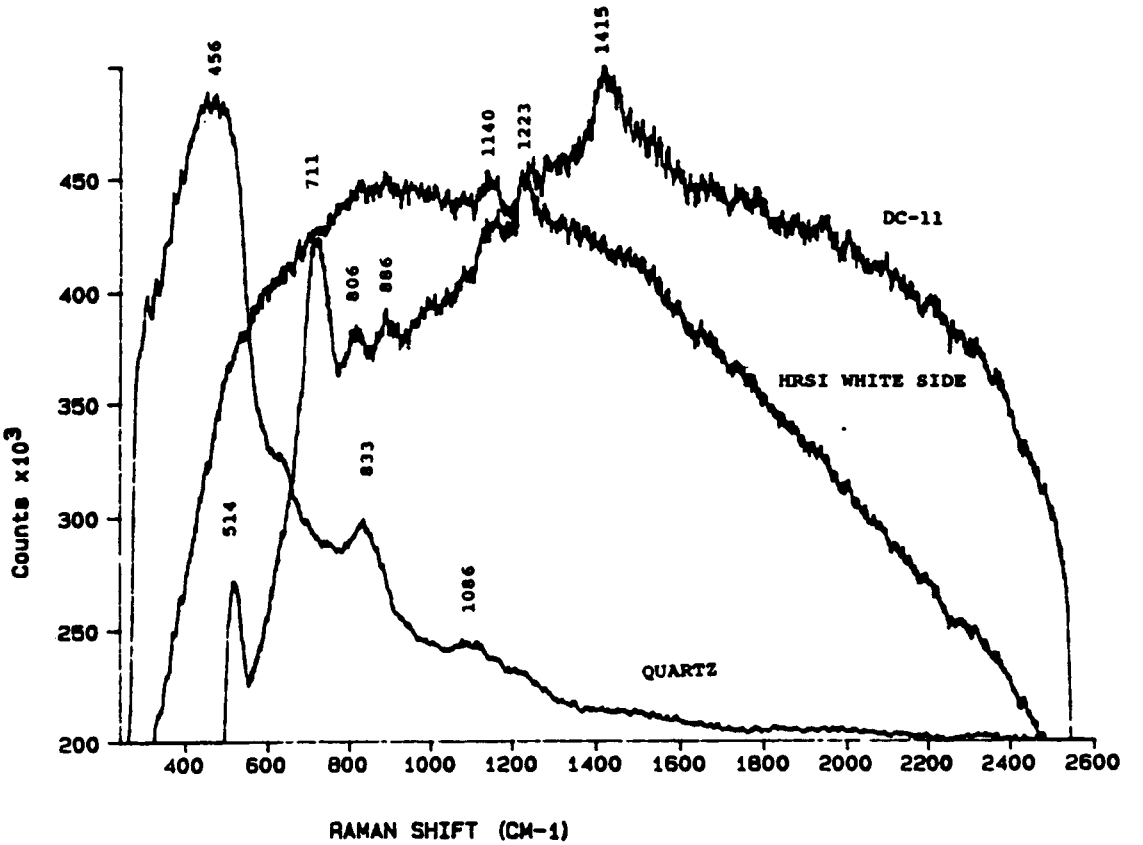


Figure 4. Raman Spectra for Various Materials, 240 to 2600 cm-1, 514.6 nm Excitation Line.

intensities are different. Further checks for Raman shifts at other laser excitation lines (488.0, 496.5, and 501.7) were made. In all cases, the peaks were found within 30 wavenumbers of those given above.

Other materials investigated for Raman shifts at room temperature are listed in Table 3. The Raman spectra for these various materials tested are shown in Figures 4, 5, & 6. All of these materials showed at least one good Raman shift or "peak" with the exception of HRSI RCG coated material. Wavenumbers of the peaks are also listed in Table 3. Full identification could not be made because of time limitations.

EMISSION AND RAMAN SPECTRA UNDER ARC JET CONDITIONS

Several experiments were conducted under arc jet conditions. Table 1, presented above, lists the conditions investigated.

Figure 7 shows the emission spectra for laser on and plasma background conditions for a hemisphere inserted into an arc jet flow of 0.0454 kg/sec of air at 400 and 500 amps. The stagnation point of the hemisphere was located 23.2 cm from the exit plane of the nozzle. The spectrometer focal point was 1.5 cm behind the stagnation point. The results showed several emission peaks. The eleven lines located between 528 to 541.1 nm are primarily related to nitrogen atoms with the 533 peak related to oxygen atoms. The broad band in the 576. nm and 583.6 nm are related to N_2 1st positive system delta v = 4. The broad peaks at 557.6 and 562.2 nm are currently not identified. More details about species identification for arc jet flow can be found in Willey, 1987.

Another example of emission spectra at a different location on the hemisphere is shown in Figure 8. The location is in a cooler region of 3.5 cm back of the stagnation point. The emission spectra is dampen somewhat. One trend that is observed is that the influence of the laser on the emission spectra is enhanced as temperature is lowered.

Figure 9 shows the difference or Raman spectra for the four conditions investigated above. Unfortunately, peaks which appear can be related back to the emission spectra. For example, the peak seen at 400 to 800 cm^{-1} are related to the atomic lines identified in Figure 7. The peak seen at 1500 cm^{-1} for the 500 amps point 1 condition appears because the difference between the valley for the laser minus background is greater than the surrounding difference seen in the neighboring peaks. Spectrometer settings for all data presented were a Spex slit opening of 600 microns, an intermediate slit opening of 2.8 mm at the 560 nm centering wavelength and 3.2 mm at the 616 nm centering wavelength.

Three emission peaks are seen at the higher wavelength range of 578 to 648 nm shown in Figure 10. The first peak at 589.3 nm is probably sodium, the second peak at 601.4 is probably 1st positive nitrogen, and the third peak at 616 nm is probably due to oxygen atoms. A repeated run is suggested at a lower Spex slit opening of 20 nm of so to check for atomic lines and molecular bands in this region.

Figure 11 is the difference spectra between the laser on and plasma background. The two peaks seen are tied to the emission peaks seen at 589.3 nm and 616 nm. An interesting observation is that the 601.4 peak did not excite as much as the other two peaks. Maybe this is a difference related to atomic versus molecular species. Other

TABLE 3. RAMAN SHIFTS OBSERVED FOR VARIOUS RAW MATERIALS AT ROOM TEMPERATURE

Material	Raman Shifts Observed (cm^{-1})	Possible Species
High Temperature Reusable Surface Insulation (HRSI)		
Reaction Cured Glass Coating (RCG)	2914 (vw)	
97% Borosilicate glass (5% B2O3,	2932 (vw)	
95% SiO2) & 3% Silicon Tetraboride	3256 (vw)	
	3497 (vw)	
White Silica/Alumina Fibers	1133 (w)	Si-O-Si
	1234 (m)	Si-O-Si
	2462 (w)	
	2629 (vw)	
	2838 (m)	
	2905 (m)	
Reinforced Carbon Carbon (RCC)		
Silicon-Carbide Coated Side	583 (w)	Si-O-Si
	630 (w)	"
	818 (s)	Si-CH3 ?
	1007 (m)	"
	1138 (w)	"
	1239 (w)	
	1480 (w)	
	2916 (s)	Si-CH3 ?
	2971 (s)	
	3123 (w)	
Uncoated RCC	--	
Tetraethylorthosilicate Coated Material (TEOS)	788 (m)	Si-?
Quartz	456 (sb)	Si-O-Si
	833 (m)	"
	1086 (w)	"
Dow Corning Silicon Grease DC-11	514 (s)	
	711 (s)	
	806 (w)	Si-O-Si
	1133 (m)	"
	1234 (m)	
	1415 (m)	
	2916 (s)	CH3? or =CH2
	2971 (s)	
	3123 (w)	

w - weak, m - moderate, s - strong, b - broad, v- very

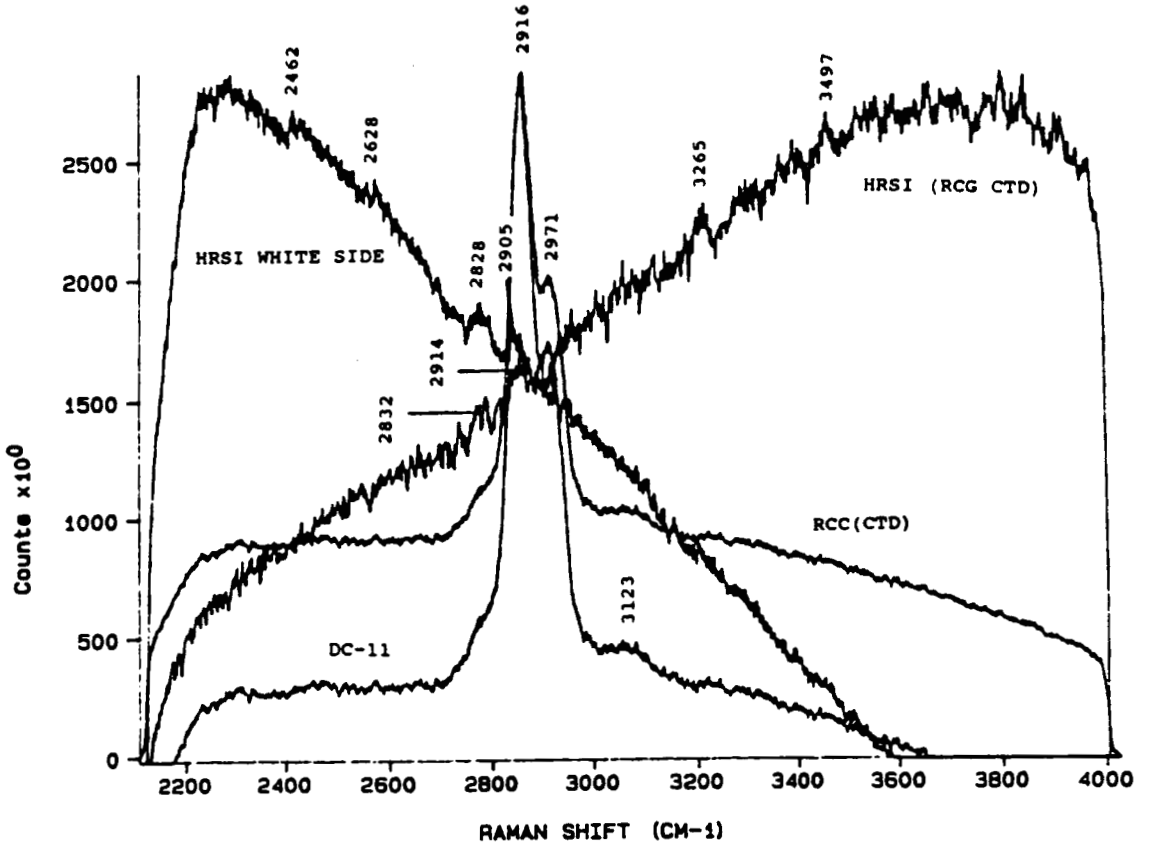


Figure 5. Raman Spectra for Various Materials, 2200 to 4000 cm-1, 514.6 nm Excitation Line.

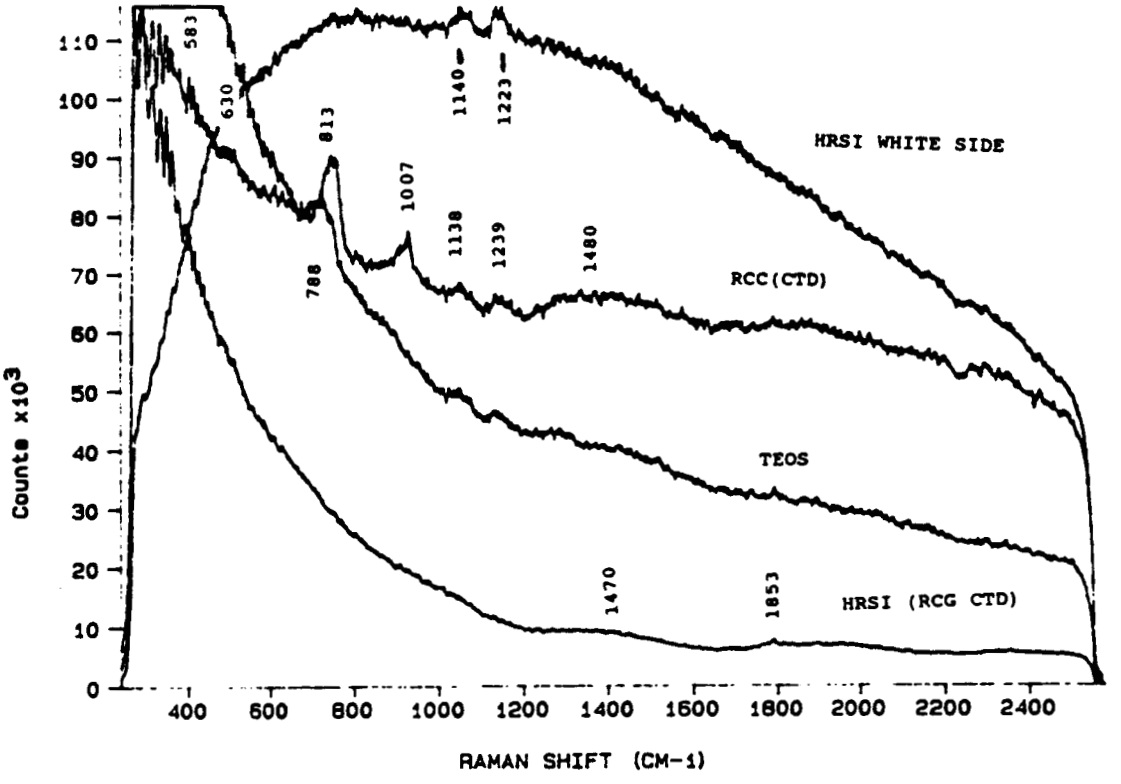


Figure 6. Raman Spectra for HRSI, RCC and TEOS materials, 240 to 2600 cm-1, 514.6 nm Excitation Line.

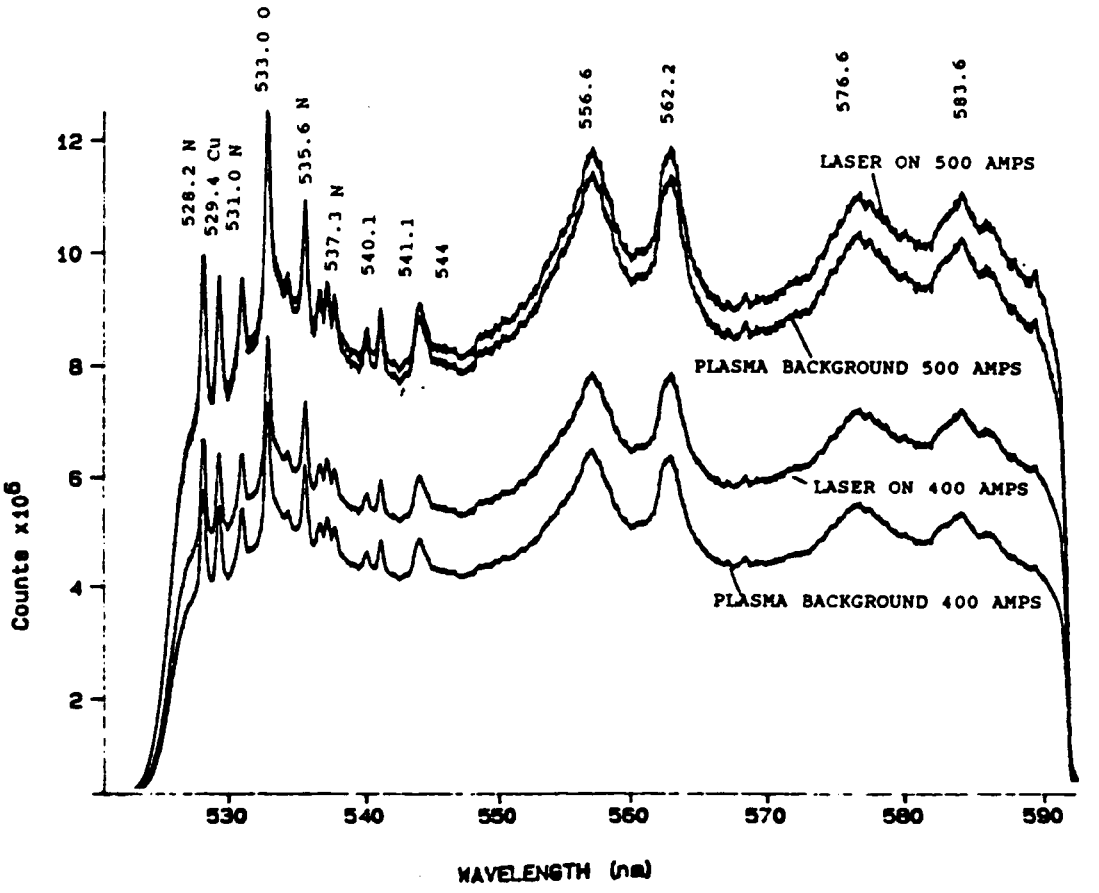


Figure 7. Emission Spectra for a HRSI Hemisphere in an Arc Jet Flow for a Point 1 cm back of the Stagnation Point. 560 nm Centering Wavelength.

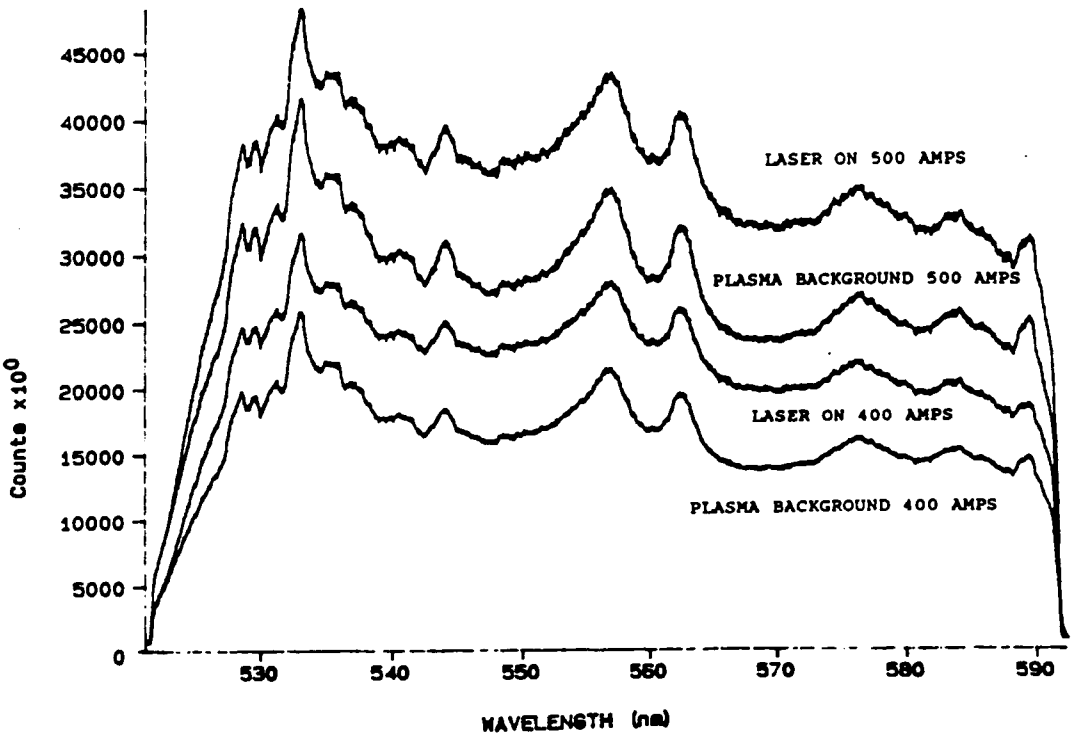


Figure 8. Emission Spectra for a HRSI Hemisphere in an Arc Jet Flow for a Point 3.5 cm back of the Stagnation Point. 560 nm Centering Wavelength.

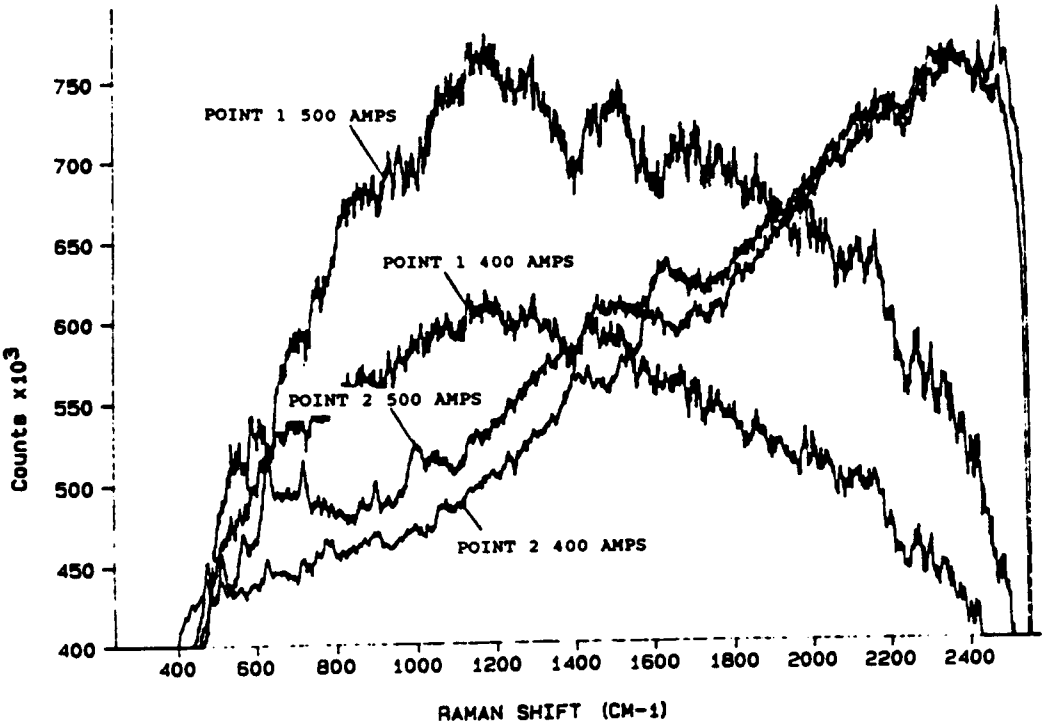


Figure 9. Raman Spectra for a HRSI Hemisphere in an Arc Jet Flow for 1 and 3.2 cm back of the Stagnation Point. 560 nm Centering Wavelength. 514.5 nm Laser Excitation Line.

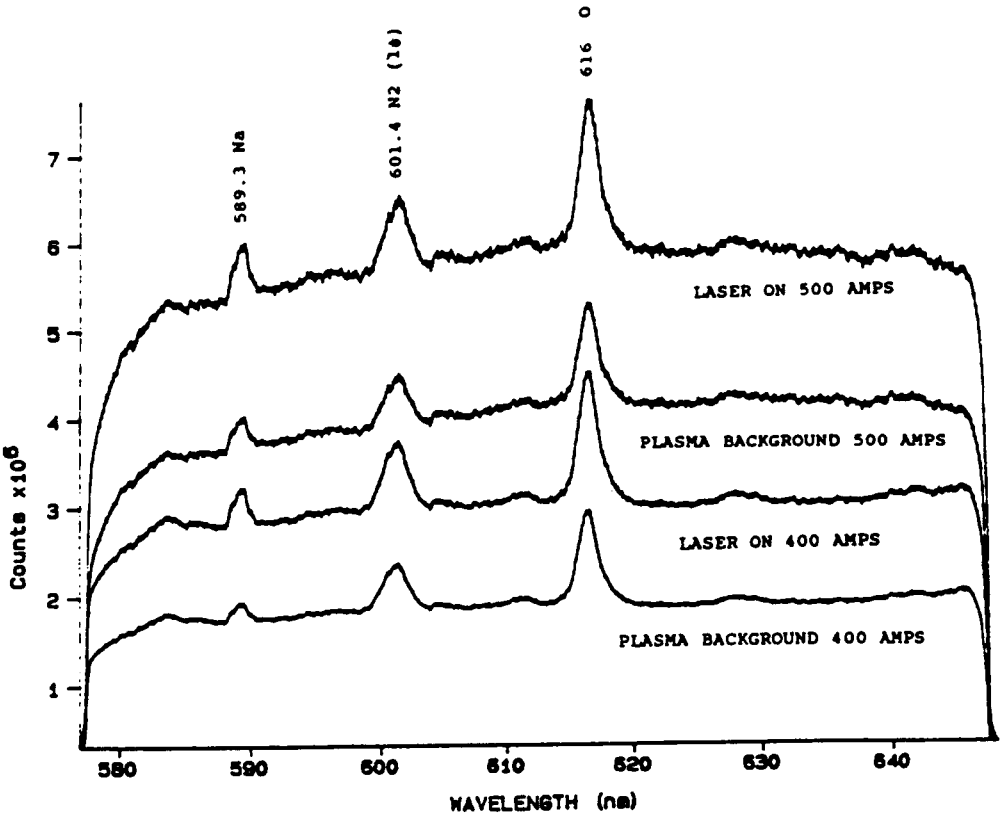


Figure 10. Emission Spectra for a HRSI Hemisphere in an Arc Jet Flow for a Point 3.5 cm back of the Stagnation Point. 616 nm Centering Wavelength.

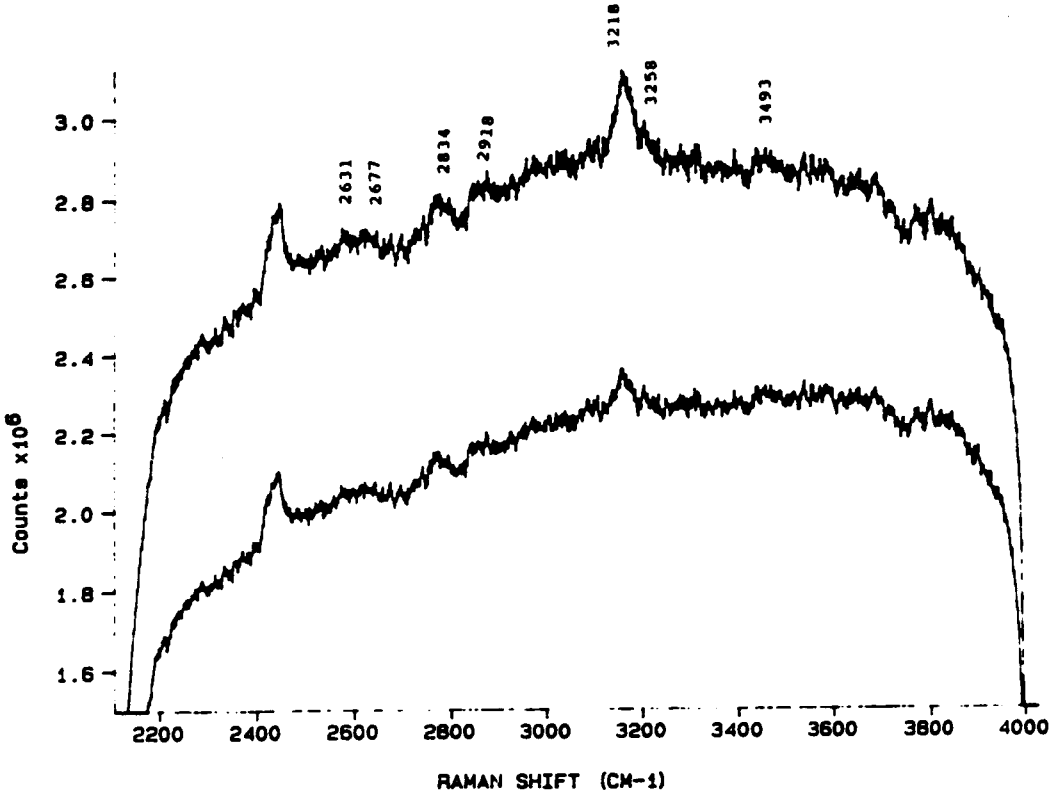


Figure 11. Raman Spectra for a HRSI Hemisphere in an Arc Jet Flow for a Point 3.5 cm back of the Stagnation Point. 616 nm Centering Wavelength. 514.5 nm Laser Excitation Line.

peaks are seen at wavenumbers of 2631, 2677, 2834, 2918, 3258, and 3493 cm^{-1}. Of these peaks, three (2631, 2677, and 3258 cm^{-1}) cannot be related back to the emission spectra and thus may represent Raman shifts. A review of the raw material Raman spectra in Figure 5 shows weak peaks at 2832, 2914, 3265, and 3497 cm^{-1}. So maybe one Raman signal may exist at 3255 to 3265 cm^{-1}. More work is suggested by isolating on this region and investigating it carefully.

Investigation on the HRSI RCG coated wedge model presented a lower temperature condition than that seen with the hemisphere. Results are shown for two current levels of 500 amps and 900 amps in Figure 12. The first noticeable result is lower intensity emission spectra for both plasma background and laser on conditions (see Figure 12). Another observation is how the two spectrums, laser on versus plasma background, differ. In the hemispherical case the laser versus plasma background spectrums were nearly identical with only intensities different. With the wedge model, this is not the case and an explanation should be found.

The difference spectra is shown in Figure 13. Bumps are present but it is difficult to determine if these are Raman shifts or noise. Further investigation is warranted to find out.

CONCLUSIONS

1. Raman spectroscopy is possible with the apparatus available at the JSC Arc Jet.

2. Raman shifts were observed under room temperature and atmospheric pressure conditions for several materials including RCC coated materials and TEOS coated materials. Weak shifts were seen for HRSI materials under the same conditions.

3. Raman shifts could not be easily discerned under arc jet conditions.

4. Laser on conditions influenced emission intensities across the regions scanned (500 to 700 nm). The influence was found to be indirectly related to arc jet temperatures. At high temperatures (1460 K) little change was observed. Where at low temperatures (980 K),major differences were seen between laser on and laser off emission spectrums.

RECOMMENDATIONS

1. Continue Raman spectroscopy investigations with HRSI materials noting the following:

   a. Confirm that the emission peaks identified are the species so identified.

   b. Repeat the Raman experiments with the Spex slit closed to 100 microns.

2. Change materials to a RCC coated material which has several good Raman peaks and repeat the Raman spectra experiments under arc jet conditions.

3. Purchase a Spex Model 1460 tunable excitation filter.

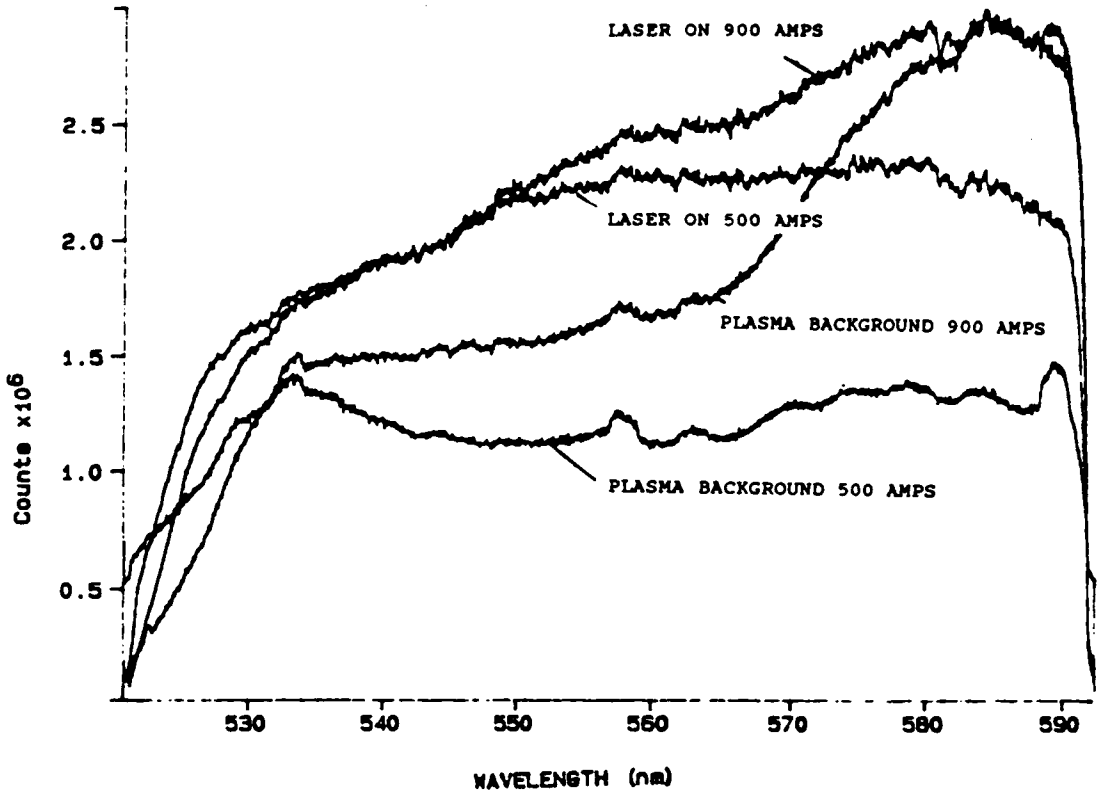


Figure 12. Emission Spectra for a HRSI Wedge in an Arc Jet Flow for a Point 3.5 cm back of the Leading Edge. 560 nm Centering Wavelength.

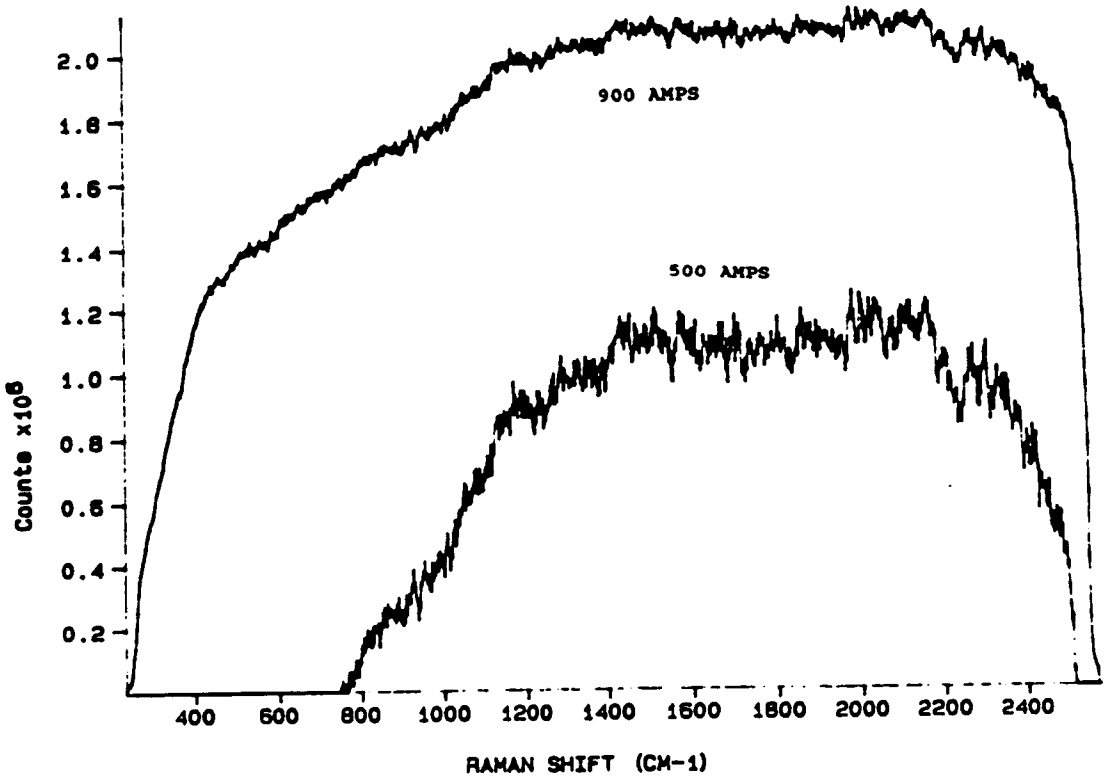


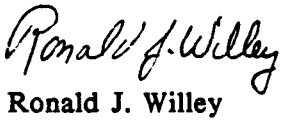
Figure 13. Raman Spectra for a HRSI Wedge in an Arc Jet Flow for a Point 3.5 cm back of the Leading Edge. 560 nm Centering Wavelength. 514.5 nm Laser Excitation Line.

4. Purchase or build a remote control for the Spex filter setting. With a plasma filter, all spectrographic experiments could be conducted inside the control room with fewer arc off requirements.

5. Continue efforts to confirm catalysis/recombination behavior at high temperature conditions. If recombination peaks with temperature do exist, heat shield design criteria may be changed substantially.

ACKNOWLEDGMENTS

I would like to express my gratitude to the many people who assisted me throughout my summer faculty fellowship. Special thanks are extended to Mr. John Grimaud, Dr. Carl Scott, Dr. Fred Wierum, Dr. Harvel Blackwell, Mr. V.S. Murty, Dr. Arepalli, Mr. Ted Heaton, and Mr. William Davis. I would also like to thank the 2nd and 3rd shift arc jet personnel for their assistance in the arc jet experiments. A thanks is extended to the Building 13 Lockheed employees, particularly Mr. Dwayne Castin who performed the infrared spectrum on the Dow Corning silicon grease. It was a pleasure to work with everyone associated with NASA and this project. THANK YOU.



Ronald J. Willey
3 September 1987

REFERENCES

Auerbach, D.R.; Tallant, D.R.; and Higgins, K.L.: Evaluation and Mapping of Heatshield Flight Temperature and Compositions with Laser-Raman Spectrocopic Techniques. AIAA-87-1638, June, 1987.

Bentley, F.F.; Bentley, F.F.; and Smithson, L.D.: Infrared Spectra and Characteristic Frequencies 700-300 cm^{-1}: A Collection of Spectra, Interpretation, and Bibliography. Interscience Publishers, Div. of John Wiley, & Sons, New York, 1968.

Blackwell, H.E.; Wierum, F.A.; and Scott, C.D.: Spectral Determination of Nitrogen Vibrational Temperatures. AIAA-87-1532, June 1987.

Chan, S.S.; and Wachs, I.E.: In Situ Laser Raman Spectroscopy of Nickel Oxide Supported on gamma -Al2O3. J. Catal., vol. 103, 1987, pp. 224-227.

Colthup, N.B.; Daly, D.H., and Wilberley, S.E.: Introduction to Infrared and Raman Spectroscopy. Academic Press, New York, 1964.

Eckbreth, A.C.; Bonczyk, P.A.; and Verdieck, J.F.: Laser Raman and Fluorescence Techniques for Practical Combustion Diagnostics. Applied Spectroscopy Reviews, vol. 13, no. 1, 1979, pp. 15-167.

Gilson, T.R.; and Hendra, P.J.: Laser Raman Spectroscopy. John Wiley & Sons, Ltd. Bath, G.B., 1970.

Greaves, J.C.; and Linett, J.W.: Recombination of Atoms at Surfaces. Advisory Group for Aeronautical Research and Development Report #332
Paris,1959.

Greaves, J.C.; and Linett, J.W.: The Recombination of Oxygens Atoms at Surfaces. Trans. of the Faraday Society, vol. 54, Sept., 1958, p. 1323.

Hair, M.L.: Infrared Spectroscopy in Surface Chemistry. Marcel Dekker, Inc., New York, 1967.

Hemley, R.J.; Bell, P.M.; and Mao, H.K.: Laser Techniques in High-Pressure Geophysics. Science, vol. 237, 1987, pp. 605-612.

Herzberg, G.: Atomic Spectra and Atomic Structure. Dover Publications, New York, 1944, p. 10.

JSC Publications Manual: Guidelines for Preparing Scientific and Technical Information for Publication in NASA Formal and Informal Documents. JSCM 2220, Mar. 1984.

Jumper, E.J.; Ultee, C.J.; Dorko, E.A.: A Model for Fluorine Atom Recombination on a Nickel Surface. J. Phys. Chem., vol. 84, 1980, pp. 41-50.

Kolodziej, P.; Stewart, D.A.: Nitrogen Recombination on High Temperature Reusable Surface Insulation and the Analysis of its Effect on Surface Catalysis. AIAA-87-1637, June, 1987.

Morrow, B.A.: Raman Spectroscopic Studies of Surface Species. in Vibrational Spectroscopies for Adsorbed Species. ACS Sym. Ser. 137 1980.

Nordine, P.C.; Fujimoto, G.T.; and Greene, F.T.: The Study of Excited Oxygen Molecule Gas Species Production and Quenching on Thermal Protection System Materials. Final Report NASA Contract No. NAS9-17261, June 24, 1987.

Pearse, R.W.B.; and Gaydon, A.G.: The Identification of Molecular Spectra. Chapman & Hall, London, 1965.

Rochelle, W.C.; Battley, H.H.; Grimaud, J.E., Tillian, D.J.; Murray, L.P.; Lueke, W.J.; and Heaton, T.M.: Obiter TPS Development and Certification Testing At the NASA/JSC 10 MW Atmospheric Reentry Materials and Structures Evaluation Facility. AIAA-83-0147, Jan 1983.

Scott, C.D.: Effect of Nonequilibrium and Wall Catalysis on Shuttle Heat Transfer. Journal of Spacecraft and Rockets, vol. 22, no. 5, Sept-Oct, 1985, pp. 489-499.

Seward, W.A.: A Model for Oxygen Atom Recombination on a Silicon Dioxide Surface. AFIT/DS/AA/85-1, Ph.D. Dissertation, Air Force Institute of Technology, 1985.

Stewart, D.A.; and Leiser, D.B.: Catalytic Surface Effect on Ceramic Coatings for an Aeroassisted Orbital Transfer Vehicle. Ceramic Engineering and Science Proceedings, vol. 5, March, 1984, pp. 491-505.

The Encyclopedia Britannica: Encyclopedia Britannica, Inc. Chicago, 15th Ed., 1985, pp. 9:917, 13:554-581.

Wachs, I.E.; Hardcastle, F.D.; and Chan, S.S.: Raman Spectroscopy of Supported Metal Oxide Catalysts. Spectroscopy, vol. 1, no. 8, 1986.

Wierum, F.A.: 1984 ASEE-NASA Summer Faculty Fellowship Final Report. NASA-JSC 1984.

Willey, R.J.: The Identification of Excited Species in Arc Jet flow. NASA CR 171990, August 15, 1987.

Wise, H.; Wood, B.J.: Reactive Collisions Between Gas and Surface Atoms. Adv. in Atomic & Molecular Physics, vol. 3, 1967, pp. 291-353.

FRACTURE STUDY OF WINDSHIELD GLASS PANES

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	H. Y. Yeh
Academic Rank:	Associate Professor
University & Department:	Prairie View A&M UNiversity Civil Engineering Department Prairie View, Texas 77446
NASA/JSC	
Directorate:	Engineering
Division:	Structures & Mechanics
Branch:	Structural Mechanics
JSC Colleague:	Orvis Pigg
Date:	August 14, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

The major stresses to cause crack propagation in windshield glass panes are induced by bending moment which is resulted from the pressure diffierentials across the panes. Hence the stress intensity factors for finite plate with semi-elliptical surface flaw and edge crack under bending moments are examined. The results show that the crack growth will be upperbound if it is computed by using the stress intensity factor for finite plate with edge crack subjected to pure bending moments. Furthermore, if the ratio of crack depth to plate thickness, a/t, is within 0.3, the stress intensity factor can be conservatively assumed to be constant of being the value at a/t equal to zero. A simplified equation to predict structural life of glass panes is derived based on constant stress intensity factor. The accuracy of structural life is mainly dependent on how close the empirical parameter, m, can be estimated.

Introduction

Most structural materials contain certain surface flaw which is either inherent in the basic materials or is introduced during manufacturing, assemblying or transporting processes. The surface flaw in glass panes is one of the major factors that affect the glass strength[1]**. The flaw depth of the glass is controlled by the way how the surface is polished. Its depth is usually at least three times of the diameter of the grinding particles used in surfacing finishing operation[2]. If the grinding techniques can be improved to a degree so that no surface flaw exits, the strength of the glass can reach as high as 2,000,000 psi. Without any doubt, the structural life of glass pane can be prolonged by reducing the surface flaws. Nevertheless, today's most sophisticated grinding technology can not diminish the flaw depth beyond 0.001 inch. The inner glass panes of flight vehicles are constantly subjected to pressure differentials across the panes. Consequently, the incipient flaw existing in the glass panes will start to propagate because of the flexural stresses inside the glass induced by pressures. The elastic theory of fracture mechanics has been widely utilized to investigate the induced stress in the crack zone and the behaviour of crack propagation of the glasses. Since the states of elastic stress and strain in the vicinity of the crack are characterized by the stress-intensity factors, extensive research in this area has been conducted in the past two decades. Exact solutions as well as approximate solutions for various flaw shapes in infinite as well as finite bodies have been obtained[3,4].

** Numerals in the brackets refer to the list of references.

The objectives of this study are to investigate the stress-intensity factors for analyzing the crack propagation of glass panes and to seek the effecient way of predicting the structural life glass of panes under service loads.

Stress-Intensity Factors

Fracture mechanics undoubtly has been accepted by most design engineer as a major tool to prevent the structural system from brittle failure. Since imperfection exists in most engineering materials, attention has been concentrated on analyzing the stress redistribution in the small region of flaw to ensure the material toughness in flaw region is strong enough to avoid failure. This toughness is characterized as fracture toughness K_{Ic}, which is a constant for given materials. The crack size corresponding to fracture toughness for the specified stress is termed as critical crack size. In case the crack size passes beyond the critical threshold, crack propagation becomes unstable and fracture occurs. The stress-intensity facor was introduced to describe the stress behaviour in the neighborhood of crack tip for the crack size smaller than critical size. Therefore the main subject of fracture mechanics is to obtain the solution of stress-intensity factors for different problems.

In 1921, Griffith[5] uses the energy approach to establish the basic equation of stress-intensity factor for an infinite cracked flate sheet with a central crack subjected to a remote uniform tensile stress which is perpendicular to the crack. Griffith's formula only applies to infinite bodies. Errors may arise due to the finite size of crack solid as well as the plasticity

effect. Consequently, considerable efforts have been extended to investigate the stress-intensity factors for finite solids. Because of the complexities of the problems, the eaxct solutions of stress-intensity factors for various crack configurations in finite bodies are not available. Numerical techniques, such as finite element, boundary collocation, mapping, intergral transform, and asymptotic expansion, etc. have been extensively used by many reseachers to seek the approximate solutions for different cases. Shih[3] and rooke[4] sumarize the stress-intensity factors for various configurations in great details.

Since the semi-elliptical surface flaw has been recognized as a closed approximation to natural flaw and since the major force to drive the crack propagation in glass panes is flexural stresses, it is of much interest to assess the effectiveness of the various stress-intensity factors of elliptical surface flaw in measureing crack growth in glasses.

The stress-intensity factors for a shallow, semi-elliptical surface flaw in a plate with finite thickness under uniform tension was developed by Irwin[6] basing upon Green and Sueddon [7] solution. Method of superposition and iteration has been used to find the solution for semi-circular surface flaw in half space under tension, bending load or other loading combination[8,9]. Smith[10] obtained an approximate solution for a semi-elliptical surface flaw in a plate with finite thickness due to tension and bending moment. Finite element method[11,12,13,14] and boundary-interal equation[15] also have been applied to such complex problems of two or three dimensional solids. Since no exact solution exists, experimental studies were conducted to assess

the accuracy of the approximate solution derived from numerical techniques[16,17,18].

The solutions of stress-intensity factors of the edge crack in a plate with finite thickness were investigated by Gross et.al.[19,20] using boundary collocation method. Their solutions agree well with the experimental data[21,22].

The solution of stress-intensity factor of semi-elliptical surface flaw in a plate with finite width can be expressed in the following general form;

$$K_I = M S \sqrt{\pi a}/Q \qquad \text{----------(1)}$$

where $Q = \phi^2 - 0.212 (S_o/S_y)^2$

and $\phi = \int_0^{\pi/2} \{ \cos^2 \emptyset + (a/c)^2 \sin^2 \emptyset \}^{1/2} d\emptyset$

K_I is the stress intensity factor; S is the maxmium flexural stress at the outer fiber; S_y is the yield strength and a and c are the minor and major axes of the ellipse, respectively. M is a dimensionless magnification factor, dependent upon the crack geometry, crack depth, plate thickness and stress location. The second term in function Q is the correction factor for plasticity effect and will be neglected in this study since the glass is a brittle material. Q is equal to one for edge crack as a/c approches zero and becomes $\pi/2$ for semi-circular crack as a = c. In order to visualize the effect of parameters on the value of magnification factor M, Smith's[10], Marrs'[17] and Gross'[20,22] solutions at the point of maximum crack depth for surface flaw are plotted in Fig. 1 as function of a/c and a/t. When the ratio a/c approaches zero, Smith's solution closely agrees with Gross'

solution, which can be expressed in terms of power series as follows;

$$M = 1.12 - 1.39(a/t) + 7.32(a/t)^2 - 13.1(a/t)^3 + 14(a/t)^4 \quad \text{----(2a)}$$

The assumptions used to derive equation 2a are (1) finite thickness plate with central edge crack subjected to pure bending moments; (2) the direction of moment is perpendicular to crack plane; and (3) no surface traction exists on both front and back surfaces. Also Equation 2a is only valid for the ratio of crack depth to plate thickness not greater than 0.6. Within this limit, the error by using Equation 2a is less than 1%. In case shear force and bending moment exert simultaneously to the plate, the stress intensity factors are influenced by the ratio of plate span to plate thickness. The magnification factors calculated by Gross[22] for a simply supported plate with a span of L and edge crack in the central under central loads for L/t=8 and L/t=4 are shown in Equations 2b and 2c and plotted in Fig. 2.

$$M = 1.11 - 1.55(a/t) + 7.71(a/t)^2 - 13.5(a/t)^3 + 14.2(a/t)^4 \quad \text{---(2b)}$$

for $L/t = 8$

$$M = 1.09 - 1.73(a/t) + 8.20(a/t)^2 - 14.2(a/t)^3 + 14.6(a/t)^4 \quad \text{---(2c)}$$

for $L/t = 4$

As the span is lengthened, the effect of shear force becomes negligible. Hence the magnification factor for the three-point moment approaches to that of pure bending as it is illustrated in Fig. 2. Fig. 1 or 2 explains that the magnification factors decrease, at first, as a/t increases. Then they reach the minimum where a/t is approximately equal to 0.13 for pure bending and a/t=0.14 and 0.15 for L/t=8 and 4, respectively, for three point moment. Thereafter, both the magnification factors and a/t increase

simultaneously. Conclusion can be drawn from Figs. 1 and 2 that the stress-intensity factor of edge crack in a plate with finite thickness subjected to pure bending loads is more critical than that of semi-elliptical surface flaw under same conditions and is also more severe than that of edge crack under three-point moment. Therefore, to study the crack growth and to predict the structural life of glass panes, the stress-intensity factor of edge crack under pure bending moments should be considered to warrant the solutions in the safe domain.

Crack Growth

According to the experimental study done by Wiederhorn et.al. [23 to 28] the crack velocity of flaw propagation in glass is a function of stress-intensity factor, K_I, and the environments. In this study, the effect of the enviornments on crack velocity will be neglected. Consequently, the relationship between crack velocity and the stress-intensity factor can be expressed either in power function

$$V = D K_I^m \qquad \text{----------(3a)}$$

or in exponential form

$$V = V_o \exp[B K_I] \qquad \text{----------(3b)}$$

where D, V, B and m are empirical constants. These constants can be determined from experimental data collected by conducting double-cantiliver-beam experiments. If the flaw depth at any time, T, is denoted by a, then the crack velocity is defined as follows:

$$V = \frac{da}{dT} \qquad \text{----------(4a)}$$

or

$$da = V\, dT \qquad \text{----------(4b)}$$

Substitute Equations 1 and 2a into Equations 4b, we obtain

$$\int_{a_o}^{a_c} M^{-m}\, a^{(-m/2)}\, da = \int_{0}^{T} (D\, S\, /\, \sqrt{\pi}\,)^{m}\, dT \qquad \text{--------(5a)}$$

or

$$\int_{a_o}^{a_c} \mathrm{Exp}[B\, M\, S\, /\, \sqrt{\pi a}\,]\, da = \int_{0}^{T} V_o\, dT \qquad \text{----------(5b)}$$

Since the magnification factor is not a linear function of flaw depth, no explicit expression for crack growth can be obtained. However, Equations 5a or 5b can be solved by numerical method. In order to view the effect of parameter, a/t, on the crack propagation, a load spectrum shown in Figs. 4 is utilized to compute the crack growth from Equation 5b. The results show that the total crack growth mainly depends on stress intensity factor. As stress intensity factor decreases, the crack growth slows down. The growth rate approaches the minmum when stress intensity factor reaches the smallest value. Then the growth rate accelerates as stress intensity factor increases.

To study the fracture behaviour of the windshield glass for shuttle orbiter, Hayashida et.al. [32] used M=1.12, which corresponding to M value by setting a/t = 0 in Equation 2a. According to the graphic indication from Fig. 5, the calculated crack growth with M = 1.12 exceeds the actual crack growth by considering M as function of a/t as long as a/t is less than 0.3. Accordingly, if M is considered to be 1.12 for a/t being less than 0.3, the crack growth will be the upper bound. If M is assumed to be constant, Equation 5a can be integrated.

$$a_j = \{\, a_{j-1}^{(1-m/2)} + (1-m/2)\, D\, (S_j\, M\, /\, \sqrt{\pi}\,)^{m}\, \Delta T_j \,\}^{(1-m/2)} \qquad \text{-----(6)}$$

where S_j is the applied stress at time interval ΔT_j and a_j

is the total crack depth at time T_j. In the case, the sustained stress is constant, i.e. $S_j = S$ for all j, Equation 6 becomes

$$a_f = \{ a_o^{(1-m/2)} + (1-m/2)\, D\, (S\, M\, /\, \sqrt{\pi})^m\, T^{(1-m/2)} \} \qquad \text{-------(7)}$$

a_f is the total crack depth and T is the time duration that the sustained stess is acting upon the object.

Structural Life of Glass Panes

To study the structural life of glass panes, it is assumed that the magnification factor, M, be constant. It is justified since the variation of a/t during the crack propagation is very small. Wiederhorn et.al.[24,25] have derived the expression for predictiing the structural life of glass for skylab windows by using Equations 3b and 4a.

$$T = \frac{2\exp[-BSM/\sqrt{\pi a_o}](1+BSM/\sqrt{\pi a_o})}{(BSM/\sqrt{\pi})^2 V} \{1-\exp[-BSM/\sqrt{\pi a_o}(\frac{\sqrt{a_c}}{\sqrt{a_o}} -1)]\frac{(1+BSM/\sqrt{\pi a_c})}{(1+BSM/\sqrt{\pi a_o})}\} \quad -(8)$$

a_c is the critical flaw depth which can be determined from the fracture toughness, K_{Ic};

$$a_c = \{K_{Ic} /(SM/\sqrt{\pi})\}^2 \qquad \text{----------(9)}$$

Since, in practice, a_c is usually much larger than the initial flaw depth, a_o, after discarding small terms, Equation 8 becomes

$$T = \frac{2\exp[-BSM/\sqrt{\pi a_o}]\sqrt{a_o}}{BSM/\sqrt{\pi}\, V_o} \qquad \text{----------(10)}$$

Equation 10 can be used to estimate the structural life of glass pane if the material constants of the particular glass are known. If Equation 10 is expressed in terms of stress-intensity factor and crack velocity, it becomes

$$T = \frac{2\,a_o}{V_1\,B\,K_{I1}} \qquad \text{----------(11)}$$

V_1 is the initial crack velocity corresponding to K_{I1}.
This equation can be used to estimate the service life of glass panes if the information of the similar glass panes is given. This information can be obtained from experimental work. The service lives of the glass panes are related as follows:

$$T_2 = \frac{V_1\,K_{I1}}{V_2\,K_{I2}}\,T_1 \qquad \text{----------(12)}$$

Wiederhorn [25] has shown that the results computed from Equation 9 agree with his experimental data.

Since, as mentioned previously, the crack velocity can also be expressed in power formation, the time for flaw depth to reach the critical point is derived from Equations 3a and 4a.

$$T = \frac{2\,a_o}{D(S\,M\,\sqrt{\pi a_o})^m (m-2)}\left\{1 - \left(\frac{a_o}{a_c}\right)^{m/2-1}\right\} \qquad \text{----------(13)}$$

The last term in the right-hand side of Equation 13 can be mitted, since a_c is usually much larger than a_o.
Equation 13 can be simplified as shown in the following:

$$T = \frac{2\,a_o}{D(S\,M\,\sqrt{\pi a_o})^m (m-2)} \qquad \text{----------(14)}$$

It is to be noted that if the ratio, a_o/a_c, is not very small, the error of T calculated from Equation 11 will mainly rely on m. The larger m is, the smaller the error will be. For example, if a_o/a_c is equal to 0.5, the percentages of error of T are about 0.8% and 0.006% with m being 16 and 30,

respectively. After multiplying both sides of Equation 14 by S^m, take logarithm on both sides.

$$\ln(T) + m \ln(S) = \ln(C) \quad \text{----------(15)}$$

where $$C = \frac{2 a_o}{D(M/\sqrt{\pi a_o})^m (m-2)} \quad \text{----------(16)}$$

If they are plotted in log-log scale as shown in Fig. 6, it is a straight line with a slope of -1/m for time,T, and stress, S. Because the constant, C, is a function of the initial flaw depth, the lines for various a_o should parallel to one another. From Figure 6, the approximate life of glass under any stress level can easily be determined. The accuracy of the results will mainly be controlled by empirical constants, such as D and m.

In order to predict the service life of glass by using Equation 16, it is necessary to measure the initial flaw depth, in addition to the determination of empirical constants D and m. However, the relationship of the service lives of two identical glass panes in the same environment can be formulated from Equation 16.

$$T_2 = \left(\frac{S_1}{S_2}\right)^m \left(\frac{a_{o1}}{a_{o2}}\right)^{m/2-1} T_1 \quad \text{----------(17)}$$

If the initial flaw depths for both glass panes are the same, then Equation 17 becomes

$$T_2 = \left(\frac{S_1}{S_2}\right)^m T_1 \quad \text{----------(18)}$$

In supporting the validity of Equation 12, Wiederhorn[25]

obtained the following data from his experimental study.

$S_1 = 8.4\ MN/m^2$, $V_1 = 1.4 \times 10^{-4}\ m/s$, $T_1 = 5\ sec$

$K_{I1} = 0.7\ MN/m^{3/2}$

$S_2 = 4.2\ MN/m^2$, $V_2 = 2 \times 10^{-9}\ m/s$, $K_{I2} = 0.35\ MN/m^{3/2}$

By using Equation 12, T_2 is equal to 7×10^5 sec, which confirms with his experimental data. Though m is not available from his study, it can be estimated by substituting all V's and K 's into Equation 3a. We obtain

$$m = \frac{\ln(V_1/V_2)}{\ln(K_{I1}/K_{I2})} = 16.095 \qquad \text{----------(19)}$$

By substituting m into Equation 18, T_2 is approximately equal to 3.5×10^5 sec. It does not agree well with the experimental results due to the lack of accurate information of constant m. From Wiederhorn's study[26] of crack velocity for various silica glasses, he shows that the constant m is in the range of the 30's. Since the service time calculated from Equation 18 is very sensitive to the constant m, extensive experimental study on m is a must. If accuracy of m is reachable, the service life of glass pane computed from Equation 14 will be very accurate for large cracks. Even for a small crack, Equation 18 is still a good estimator.

Conclusions

The driving forces to initiate the propagation of surface flaw in the inner glass panes of flight vehicles are mainly due to flexural stresses induced by the pressure differentials

across the panes. Hence the stress intensity factors for various surface flaw shapes in a finite thickness plate subjected to bending loads are examined. The results show that the solutions obtained by using the stress intensity factor for plate with edge crack under pure bending moment to analyze the facture of glass panes are more conservative than that of semi-elliptical surface flaw under bending loads or edge crack under three point moment. If the ratio of the crack depth to plate thickness is within 0.3, the solution by using magnification factor equal to 1.12 will be the upper bound for crack growth and lower bound for structural life. The design based on this assumption of M being equal to 1.12 is conservative.

A simple equation (Equation 18) is obtained to predict the structural life of one glass pane based on the information of the other. Since there is no experimental data available to measure the effectiveness of Equation 18 and the result obtained from this equation is very sensitive to the value of m, it is recommended that an extensive experiment should be conducted to study the variations of parameters as well as the verification of Equation 18.

Acknowledgements

The author gratefully appreciates Mr. Orvis Pigg's interest and encouragement in this study.

References

1. Stoll, R., Forman, P., and Edelman, J., "The Effect of Different Grinding Procedures on The Strength of Scratched and Unscratched Fused Silica", presented at Symposium on the Strength of Glass and Ways to Improve It by union Scientifique Continentale du Verre, Florence, Italy, Sept. 1961.

2. Jones, F. S., "Laten Milling Marks on Glass", Journal of American Ceramic Society, Vol XXIX, 1946, pp 108

3. Shih, G. C., "Handbook of Stress-Intensity Factoes for Researches and Engineers", Institute of Fracture and Solid Mechanics, Leghigh University, Bethlehem, Pennsylvania, 1973.

4. Rooke, D. P. and Cartwright, D. J., "Compendium of Stress-Intensity Factors", Her Majesty's Stationary Office, London, 1976

5. Griffith, A. A., "The Phenomena of Rupture and Flow in Solids"' Phil. Trans. Roy. Soc. of Lodon, A221, 1921, pp 163-197

6. Irwin, G. R., "Crack Extension Force for A Part-Through Crack in A plate", Journal of Applied Mechanics, Vol. 29, Transaction of ASME, Vol. 84, Series E, Dec. 1962, pp 651-654.

7. Green, A. E. and Sneddom, I. N., "The distribution of Stress in the Neighborhood of a Flat Elliptical Crack in An Elastic Solid", Proceedings, Cambridge Philosophical Society, Vol. 46, 1950 pp. 159-163

8. Smith, F. W., Emery, A. F. and Kobayashi, A. S., "Stress-Intensity Factors for Semi-circular Cracks, Part 2- Semi-Infinite Solid", Journal of Applied Mechanics, Vol. 34, No. 4, Transactions of ASME, Series E, Vol. 89, Dec. 1967, pp 953-959.

9. Kobayashi, A. S., "CRack Displacement in A Surface Flaw Plate Subjected to Tension or Plate Bending", Proceedings of Second International Conference on Mechanical Behaviour of Materials, ASM, 1976, pp. 1073-1077.

10. Smith. F. W., "Stress Intensity Factors for A Semi-Elliptical Surface Flaw in Plate", Structural Development Research Memo., The Boeing Company, 1966.

11. Raju, I. S. and Newman, J. C. Jr., "Stress-intensity Factors for A Wide Range of Semi-Elliptical Surface Cracks in Finite-Thickness Plates", Journal of Engineering Fracture Mechanics, Vol. 11, No. 4, 1979, pp. 817-829.

12. Newman, J. C. Jr. and Raju, I. S., "Analysis of Surface Cracks in Finite Plates Under Tension or Bending Loads", NASA TP-1578,

Dec. 1979.

13. Heliot, J., Labbous, R. and Pellissier, t., "A Benchmark Problem No. 1- Semi-Elliptical Surface Crack", International Journal of Fracture, Vol. 15, No. 6, Dec. 1976, pp. R197-R202.

14. Newman, J. C. Jr., and Raju, I. S., "Stress-Intensity Factors for Cracks in Three-Dimensional Finite Bodies Subjected to Tension and Bending Loads", NASA, TM-85793, April 1984.

15. Nishioka, T. and Alfuri, S. N., "Analytical Solution for Embedded Elliptical Cracks and Finite Element-Alternating Method for Elliptical Surface Cracks Subjected to Arbitrary Loadings", Engineering Fracture Mechanics, Vol. 17, 1983, pp. 247-268.

16. Shah, R. C. and Kobayashi, A. S., "Stress Intensity Factor for An Elliptical Crack Approaching the Surface of A Plate in Bending", Proceedings of the 1971 National Symposium on Fracture Mechanics, STP 513, ASME, 1972 pp. 3-21.

17. Marrs, G. R. and Smith, C. W., "A Study of Local Stresses Near Surface Flaw in Bending Fields", Proceedings of 1971 National Symposium on Fracture Mechanics, STP 513, ASME, 1972, pp. 22-36.

18. Grandt, A. F. Jr. and Sinclair, G. M., "Stress Intensity Factors for Surface Cracks in Bending", Proceedings of 1971 National Symposium on Fracture Mechanics, STP 513, ASME, 1972, pp. 37-58.

19. Gross, B., Srawley, J. E. and Brown, W. F. Jr., "Stress-intensity Factors for A single-Edge-Notched Tension Specimen by Boundary Collocation of A Stress Function", NASA TND-2395, 1964.

20. Gross, B. and Srawley, J. E., "Stress-intensity Factors for Single-Edge-Notched Specimens in Bending or Combined Bending and Tension by Boundary Collocation of A Stress Function", NASA TND-2603, 1965.

21. Srawley, J. E., Jones, M. H. and Gross, B., "Experimental Determination of The Dependence of Crack Extension Forces on Crack Length for A Single-Edge-Notched Tension Specimen", NASA TND-2396, 1964.

22. Brown, W. F. Jr. and Srawley, J. E., "Plane Strain Crack Thoughness Testing of High Strength Metallic Materials", STP 410, ASTM, 1966, pp. 1-65.

23. Wiederhorn, S. M., "Influence of Water Vapor on Crack Propagation in Soda-Lime Glass", Journal of the American Ceramic Society, Vol. 50, No. 8, August, 1967, pp. 407-414.

24. Wiederhorn, S. M. and Bolz, L. H., "Stress Corrosion and

Static Fatigue of Glass", Journal of the American Ceramic Society, Vol 53, No. 10, Oct. 1970, pp. 534-548.

25. Wiederhorn, S. M., and Roberts, D. E., "Fracture Mechanics Study of Skylab Windows", Institute for Materials Research, NBS NASA Pr-1-168-022, T-531-A, NBS Report 10892, May 1972.

26. Widerhorn, S. M., Evans, A. G., Fuller, E. R. and Johnson, H., "Application of Fracture Mechanics to Space-Shuttle Windows", Journal of The American Ceramic Society, Vol 157, No. 7, July 1974, pp. 319-323.

27. Widerhorn, S. M. and Johnson, H., "Effect of Electrolyte PH on Crack Propagation in Glass", Journal of The American Ceramic Society, Vol 56, No. 4, April 1973, pp. 192-197.

28. Wiederhorn, S. M., " Effect of Deutrium Oxide on Crack Growth in Soda-Lime-Silica Glass", Communication of the American Ceramic Society, Dec. 1982, pp. C202-C203.

29. Wiederhorn, S. M., Fuller, E. R. Jr., Mandel, J. and Evans, A. G., " An Error Analysis of Failure Predication Techniques Derived from Fracture Mechanics", Journal of The American Ceramic Society, Vol. 59, No. 9-10, Sep.-Oct., 1976, pp. 403-411.

30. Magida, M. B., Forrest, K. A. and Heslin, T. M., " Dynamic and Static Fatigue of A Machinable Glass Ceramic", Symposium on Methods for Assessing the Structural Reliability of Brittle Materials, ASTM STP 844, 1984, pp. 81-94.

31. Akins, A. G. and Mai, Y.-W., "Elastic and Plastic Fracture", John Wiley, 1985, pp. 503-509.

32. Hayashia, K., King,G. L., Tesinsky, J. and Witenburg, D. R., "Rational for Windshield Glass System Specification Requirements for Shuttle Orbiter", NASA CR-112209, SD 72-SH-0122, North American Rockwell, Oct. 1972.

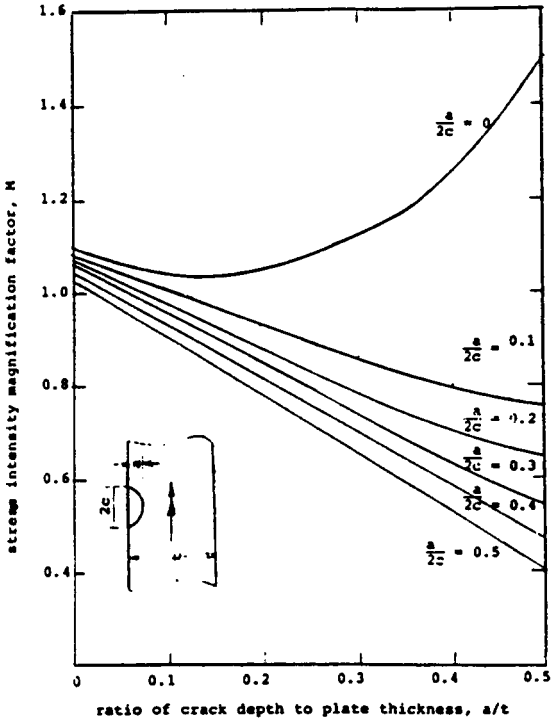


Figure 1. Stress Intensity Magnification Factor for Semi-Elliptical Surface Flaw at Maximum Depth

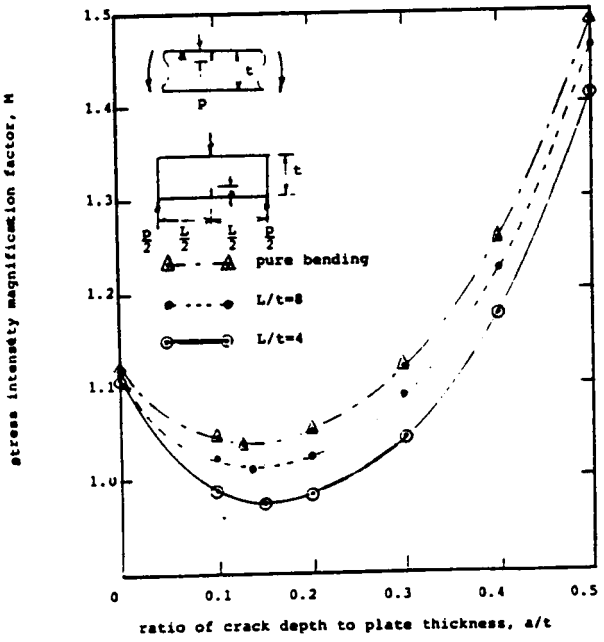


Figure 2. Stress Intensity Magnification Factor for Edged Crack Plate Under Bending Moments

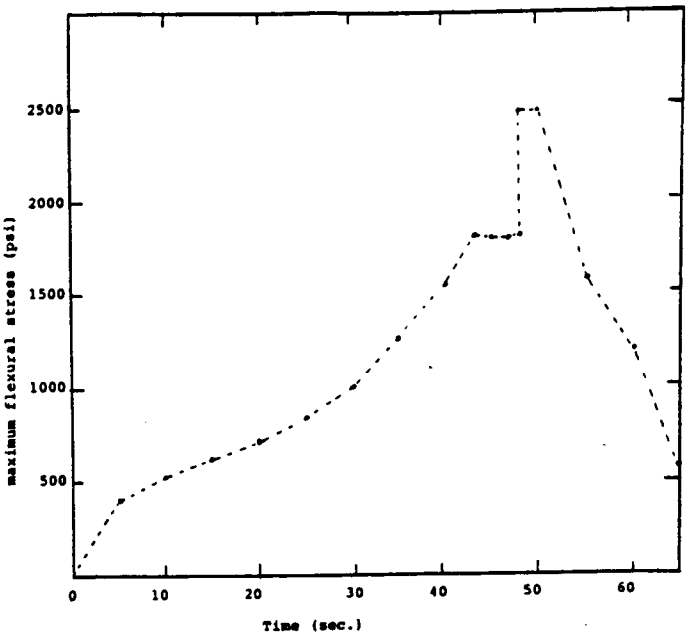


Figure 3. Load Spectrum

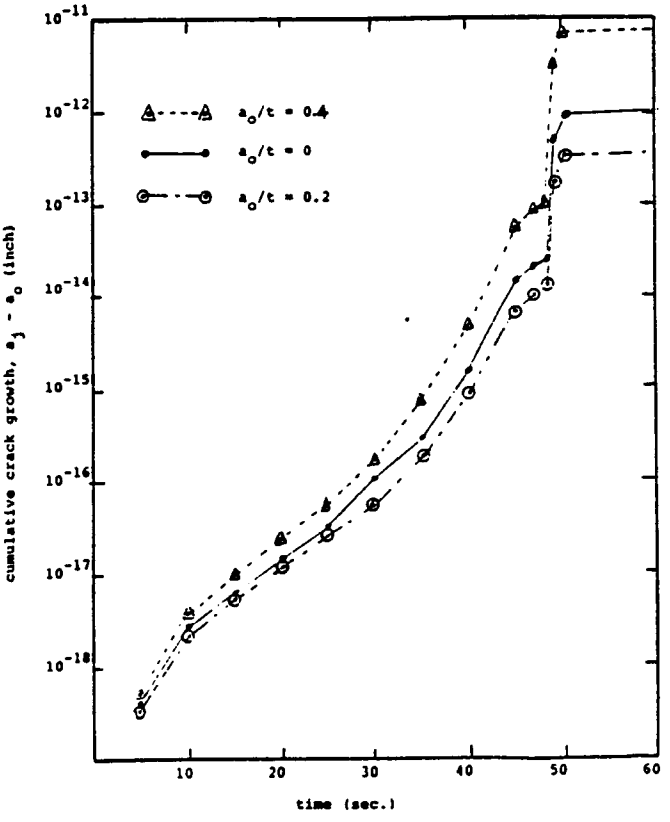


Figure 4. Cumulative Crack Growth

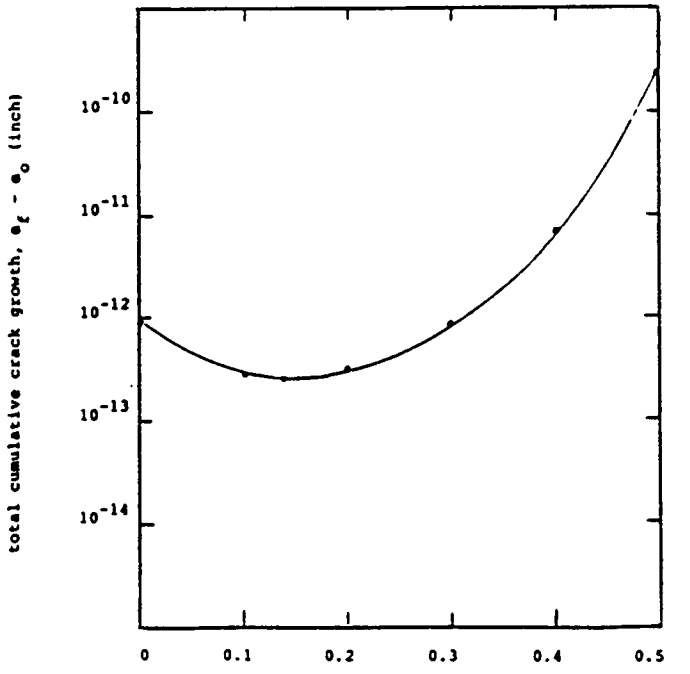


Figure 5. Total Cumulative Crack Growth

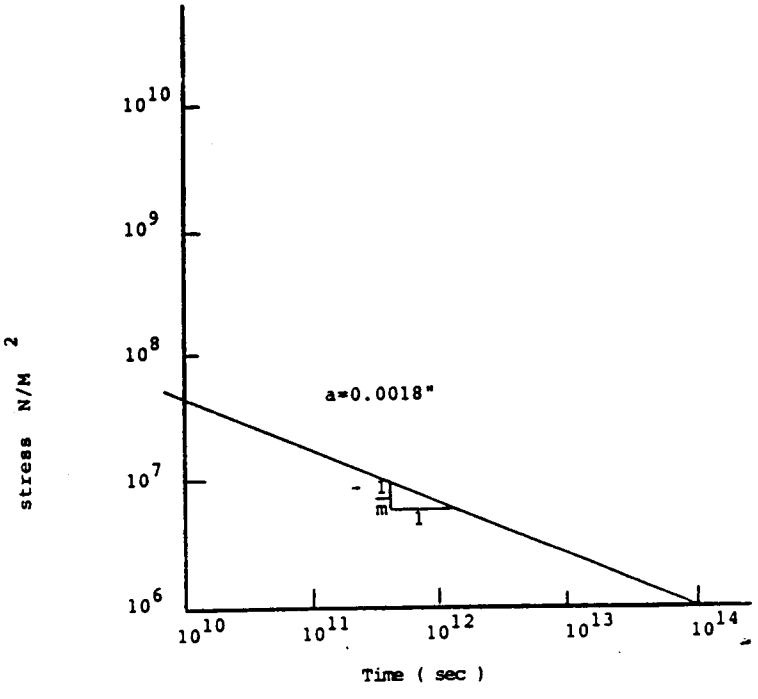


Figure 6. Stress-Time Diagram

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR 172009		

4. Title and Subtitle	5. Report Date
NASA/ASEE Summer Faculty Fellowship Program--1987 Volume 2	November 1987
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
William B. Jones, Jr. and Stanley H. Goldstein	
	10. Work Unit No.
9. Performing Organization Name and Address	
Texas A&M University and The University of Houston--University Park	11. Contract or Grant No.
	NGT-44-001-800
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The 1987 Johnson space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship program was conducted by Texas A&M University and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The basic objectives of the program, which began in 1965 at JSC and in 1964 nationally, are (a) to further the professional knowledge of qualified engineering and science faculty members; (b) to stimulate an exchange of ideas between participants and NASA; (c) to enrich and refresh the research and teaching activities of participants' institutions; and (d) to contribute to the research objectives of the NASA Centers.

Each faculty fellow spent 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1987. Volume 1 contains sections 1 through 18, and volume 2 contains sections 19 through 35.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
	Unclassified-Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price*
Unclassified	Unclassified	332	NTIS

*For sale by the National Technical Information Service, Springfield, Virginia 22161